DOCUMENT RESUME

ED 343 981 UD 028 603

AUTHOR Walker, Gary; Vilella-Velez, Frances
TITLE Anatomy of a Demonstration: The Summer Training and Education Program (STEP) from Pilot through Replication and Postprogram Impacts.
INSTITUTION Public/Private Ventures, Philadelphia, PA.
SPONS AGENCY Ford Foundation, New York, N.Y.; William and Flora Hewlett Foundation, Palo Alto, Calif.
PUB DATE 92
NOTE 83p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Daily Living Skills; Demonstration Programs; *Disadvantaged Youth; *Economically Disadvantaged; Low Income Groups; Minority Groups; Program Design; *Program Effectiveness; Program Implementation; Remedial Programs; Secondary Education; *Summer Programs; *Urban Youth; Work Experience Programs; Youth Employment; *Youth Programs
IDENTIFIERS *Summer Training and Education Program

ABSTRACT

The Summer Training and Education Program (STEP) research demonstration was initiated in 1984 to test the effects of a two-summer remediation, work, and life skills intervention on the lives of 14- and 15-year-olds from poor urban families who were already seriously behind academically. The program provided youth with half-day summer jobs combined with half-days of remedial reading and mathematics provided through the use of specially designed curricula and innovative teaching approaches. The results of research on nearly 5,000 participating youth in five cities showed that the STEP's effects were positive. Reading and mathematics test scores after the first summer were about a half grade higher for the STEP youth compared to a control group receiving only summer jobs. Participants also showed more knowledge about responsible sexual and social behavior, high attendance rates, and a high (75 percent) return rate for the second summer. However, follow-up research found that on returning to regular life and school, STEP youth experienced the same school dropout, college entrance, teenage pregnancy, and employment rates as did control group youth. This report presents the above research and findings, describes the development and implementation of the program from its beginning, and explores and analyzes the experiences of the STEP program for lessons to apply to later initiatives. Included are 11 figures and 41 references. (JB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

P/PV

Anatomy of a Demonstration

ED343981

The Summer Training and Education Program (STEP) from Pilot through Replication and Postprogram Impacts

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

N. Jaffe
Public/Private Ventures

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Gary Walker
Frances Vilella-Velez

UD 028 603

BEST COPY AVAILABLE

Rex D. Adams
Chairman of the Board

Michael A. Bailin
President

Gary Walker
Executive Vice President

Board of Directors
Rex D. Adams
Michael A. Bailin
Angela Glover Blackwell
Roy J. Bostock
Alan K. Campbell
Hillary Rodham Clinton
Alonzo A. Crim
Alice F. Emerson
Sandra Feldman
Mitchell S. Fromstein
David W. Hornbeck
Harold Howe II
Siobhan Oppenheimer-Nicolau
Marion Pines
John W. Porter
Harold Richman
Mitchell Sviridoff

Research Advisory Committee
Frank Furstenberg
Robinson Hollister
Henry Levin
Frank Levy
Anita Summers
Marta Tienda

Corporate Associates
American Express
Philanthropic Program
Ameritech Foundation
ARA Services, Inc.
Borden Foundation
BP America
Citibank
D'Arcy, Masius, Benton
and Bowles
Dean Witter Reynolds, Inc.
Exxon Corporation
Mobil Foundation, Inc.
Penn Mutual Charitable Trust
The Prudential Foundation
Rhone-Poulenc Rorer Inc.
United Technologies

Public/Private Ventures is a national, not-for-profit corporation that designs, manages and evaluates social policy initiatives aimed at helping young people whose lack of preparation for the work force hampers their chances for productive lives. P/PV's work is supported by funds from both the public and private sectors.

Acknowledgements

The Ford Foundation originated the idea that became the Summer Training and Education Program (STEP); was the major funder of its development, demonstration and research; and, with The William and Flora Hewlett Foundation, funded the long-term impact research. Other funders of STEP's demonstration and research were the U.S. Department of Labor, The Robert Wood Johnson Foundation, the U.S. Department of Health and Human Services, William T. Grant Foundation, The Edna McConnell Clark Foundation, The Hewlett Foundation, Lilly Endowment Inc., The Ahmanson Foundation, Aetna Life and Casualty Foundation, and James C. Penney Foundation, Inc.

Replication of STEP has been made possible, primarily, by support from the United States Department of Labor. Replication's other funders are General Motors, Exxon Corporation, the Rockefeller Foundation, and the states and some sites in which STEP programs are located.

Implementation of the STEP model has been the responsibility of the many local public and voluntary institutions who operate STEP in the sites. Without the interest, indeed the dedication, of the staffs of these organizations--managers, counselors, teachers--there would be no STEP story to tell.

Copyright © 1992 Public/Private Ventures

Contents

Executive Summary i

I. Introduction 1
STEP 3
The Problem and Its Consequences ... 4
Causes and Associations 6
The Intervention Strategy 8

II. Program and Research Designs ... 13
Program Model Design 13
Investment in Innovations 17
Research Design 18

III. Testing the Model 23
The Pilot 23
The Demonstration 25
Key Operating Issues 31
Early Impacts 34

IV. Replicating STEP 41
The Replication Strategy 42
Results 46
Implementation Lessons 52
Future Directions 53

V. Long-Term Impacts 57
STEP Youth in 1990 59
Why No Long-Term Impacts? 62

VI. Thoughts for the Future 67

References 73

Executive Summary

The Summer Training and Education Program (STEP) research demonstration was initiated in 1984 to test the effects of a two-summer remediation, work and life skills intervention on the lives of 14- and 15-year-olds from poor urban families who were already seriously behind academically. The program provided youth with half-days of summer jobs under the federally funded Summer Youth Employment and Training Program, combined with half-days of remedial reading and math provided by specially designed curricula and innovative teaching approaches. One half-day each week was devoted to issues concerning decision-making and responsible sexual and social behavior in the youth's everyday lives.

The research, utilizing a random assignment methodology on nearly 5,000 youth in five cities, showed that STEP's summer effects were consistently and impressively positive. Reading and math test scores after the first STEP summer were about one-half grade higher for youth receiving STEP than for a control group receiving only summer jobs. STEP youth also had significantly higher scores on tests measuring knowledge about sexually responsible social and sexual behavior. STEP youth--one-third of whom had been held back in school--had high attendance rates in the program, and a high return rate (75%) for the second summer.

The impressive summer impacts did not hold up once youth left STEP and returned to their regular school and life routines. STEP youth experienced the same school dropout, college entrance, teenage pregnancy and employment rates as the control group youth. Evidently, a positive and successful experience in work, education and life skills over two summers was not sufficient to alter the life trajectories of poor urban youth.

The problems of many of these youth--both STEP treatments and controls--persisted and grew during the follow-up period. Three and a half years after their STEP experience--at the ages of 17 and 18--15 percent were neither working nor in school; 22 percent of the young women had children and 64 percent of these mothers were receiving public assistance in some form. For youth who had enrolled in STEP a year earlier, and who were 18 and 19 at the time of the 1990 research interviews (four and a half years after leaving STEP) 25 percent were neither in school nor working; 33 percent were single parents; and 75 percent of the young mothers had received some form of public assistance. The early hope generated by STEP's summer success faded over time for large numbers of youth.

Anatomy of a Demonstration documents and analyzes the STEP initiative, from its early conceptualization and formation, through the five-city demonstration, to its replication in over 100 locations throughout the nation. Its intent is both to report hard data and to mine STEP's eight-year research and 100-site operational experience for lessons that might guide future initiatives aimed at improving the life prospects of our nation's poor youth. This mining seems especially important at the present time, when the number and proportion of youth living in poverty has been steadily increasing--now standing at more than one youth in every five--and when the nation's schools in poor areas are having great difficulty both in educating and providing other needed services to those youth.

STEP's major lessons can be summarized as follows:

1. **It is possible to produce improvements in the reading, math and life skills of young adolescents in a short period of time**. The STEP research shows unequivocally that test scores can be improved in short time periods, at reasonable operating costs. The STEP program costs $600 per youth per summer above the costs of the existing summer jobs program. The STEP experience and data counter the view that programs cannot have positive effects on teenagers who have already experienced failure in school and the debilitating effects of an environment poor in resources and opportunity.

2. **Accomplishing such improvements requires a businesslike approach to investments in innovation development**. STEP's designers and funders were willing to invest the resources necessary to produce innovative educational approaches and materials that learning theory indicated would generate improvements. To generate these resources required a consortium of philanthropic, public and corporate institutions, and an intermediary agency to organize the funding and carry out the work.

   Developing the innovations required an initial investment of about half a million dollars. Installing them cost about $15,000 per site; ongoing training, quality control and materials improvement cost another $10,000 to $15,000 per site during the initial two years of operation. These kinds of costs are familiar to business, but are often not adequately considered or taken into account in dealing with social issues. They are the costs of change.

3. **Innovative programs can be replicated with consistent practices and results across large numbers of varying locations**. Effective social programs are viewed by many as idiosyncratic and unique, dependent on exceptional local leaders and incapable of being brought to scale. The STEP experience suggests that this view may be in part a result of inadequate social investment in the packaging of substantive innovations, in the training of state and local staff in their operation, and in the use of quality-control mechanisms.

   STEP was replicated in more than 100 jurisdictions with widely varying conditions, but with consistent implementation quality and results. STEP's consistency in generating strong summer results in widely varying locations required more than the brief training sessions and standardized materials that are typically used to encourage replication. STEP's replication costs were high in comparison with the usual costs of disseminating social programs, but the replicated model got largely positive and consistent results. In addition, the STEP program has taken root in the 100 communities in which it was installed, and has expanded in a third of those communities.

4. **Short-term programs like STEP, even when they have positive short-term effects, are unlikely to produce long-lasting impacts. STEP's strength was its highly organized, intensive and controlled involvement in each youth's life.** STEP's strength was its highly organized, intensive and controlled involvement in each youth's life during the summer. But once that involvement ended, there was no vehicle to reinforce and continue STEP's positive impacts. Thus, those impacts were not carried forward, and dissipated.

STEP's major lesson for policymakers and leaders is that short-term interventions like STEP do fill critical gaps in the lives of disadvantaged young people, and do provide youth with much-needed boosts and experiences, but cannot alone produce long-term change. This does not suggest that short-term interventions have no role in a youth's development. Successful transition to adulthood depends on having a number of positive experiences as one matures, and for too many poor youth, positive experiences are few and unproductive periods are frequent. STEP's design was a response to documented evidence of significant learning loss during one such unproductive period--the summer.

STEP's record in making productive use of a critical "gap period" for an adolescent living in a poor neighborhood suggests the potential usefulness of identifying other gap periods, such as after school and weekends, and investing in the development of innovative programming that can produce short-term educational gains and life skills development. The STEP experience indicates that appropriate packaging and installation of such programming can improve the use of those gap periods in large numbers of communities.

But STEP also raises several critical questions about how more adolescents from poor families and communities can be helped to grow successfully into adulthood. One concerns the need to develop and test mechanisms and supports that can extend the impact of successful short-term interventions like STEP, which take place in gap periods. Greater involvement of families, community institutions like churches and Boys and Girls Clubs, and unrelated adults who will provide long-term guidance and support, deserve further exploration in future youth initiatives.

Another major issue is how to transfer the lessons that initiatives like STEP generate to larger youth-serving institutions, especially schools. STEP's ability to generate interest, involvement and improvements in test scores among adolescents with serious educational deficiencies suggests the need for investment in new educational approaches, and in their use in our schools. It also suggests that work and intense adult contact may be important components of a successful learning experience. The development and installation costs of such innovations--not their ongoing operational costs--are key missing ingredients in many current school reform initiatives.

The STEP long-term research also portrays dramatically the formidable obstacles that poverty presents to all efforts to improve the life trajectories of children and youth. Poverty frequently brings with it a lack of naturally occurring supports and visible opportunities, and a surfeit of negative influences and poor examples. Efforts like STEP show that it is possible to begin the process of achievement and success, even at ages 14 and 15, but that without continuity of effort, the conditions of poverty often prevail. That continuity is necessary no matter what ages a special intervention begins, so long as the youth's natural environment is poor in resources, opportunities and examples.

The STEP experience offers a complex set of lessons. They cannot be reduced to slogans, nor do they inspire easy solutions. They support neither hopelessness nor an easy optimism. They challenge our public leaders to approach the lives of an increasing number of American youth with the resourcefulness, hardheadedness, commitment and money that investment in change requires.

I. Introduction

During the past two decades, a new approach--the research demonstration--has been developed as a strategy in society's effort to improve the economic prospects of less advantaged citizens. This approach attempts to apply the practices and rigor of the laboratory in a "real-world" situation. A problem (such as dropping out of school) and its consequences (low earning ability) are selected; an intervention is designed to alleviate the problem; local organizations that can implement the intervention are identified; a sizable group of people with the problem are randomly divided into treatments (those who participate in the intervention) and controls (those who do not); the intervention program is operated; data relevant to the identified problem are collected from treatments and controls before, during and at various points after intervention operation; and sophisticated statistical procedures are applied to the data to determine whether the intervention did, in fact, alleviate the problem and its consequences.

The basic theory behind the practice and use of research demonstrations is straightforward: if the intervention "works"that is, if significantly fewer treatments than controls are affected by the problem and its consequences-- demand for the intervention should be high around the country. And that demand should generate wide-scale replication in numerous localities. The problem would thus be greatly reduced in scope. If the intervention does not work, those in the social policy field would benefit from the failure by analyzing it, then designing a new intervention with a better prospect for success.

The research demonstration approach has had strong support in part because it offers a way to temper, through the use of credible evidence, the often counterproductive partisanship that historically accompanies the consideration of social problems. It establishes a purportedly

scientific, apolitical process by which solutions can be found, and even the failures are, as in any science, important parts of the solution-building process. It also establishes the resolution of social ills as an endeavor deserving of research as rigorous as that involved in finding new vaccines, building more effective weapons or developing more fuel efficient engines. In an era when the public's trust in the political process to resolve tough social issues fairly is at a serious low point, the research demonstration offers an avenue of hope.

Two decades of experience have shown that this process of testing a social intervention can be successfully implemented. The mechanics of the research demonstration procedure are workable. People eligible for the intervention can be fairly and accurately randomized, i.e., selected by lottery to participate or not participate in the program; similar interventions can be mounted and implemented in different locations; relevant data can be collected over long periods of time for both treatments and controls; and methodologically credible results can be computed.

Although only a dozen or so large-scale, rigorous research demonstrations have been mounted, their consistency in producing results acceptable to the scientific-minded of all political persuasions has been noteworthy. The most recent example is the Manpower Demonstration Research Corporation's work in testing various approaches to moving Aid to Families with Dependent Children (AFDC) recipients off the welfare rolls and into regular employment; both political parties used the evidence produced through this demonstration in the legislative process of drafting the 1988 Family Security Act.

But at the same time, the research demonstration record has been frustrating. The interventions tested have rarely produced long-term impacts of magnitude. The strongest results in the aforementioned AFDC demonstrations did not show reductions in poverty levels or need for public assistance. **Why** the interventions tested have had such modest impact on the lives of their participants cannot be answered with the scientific precision of the impact findings themselves; thus, the partisanship that characterizes attempts to explain why some people succeed and others fail in our economy has not been substantially reduced.

In addition, even for those interventions that have shown positive impacts, replication in other localities has been minimal. Thus, the wider use of a successful intervention--a powerful justification for the research demonstration--has rarely occurred. Whether this is due more to the market structure--the financing and incentives--of the social service world, the capacity of the agencies who administer and operate such interventions, flaws in the products of the research demonstration process, or other factors is an issue that, until recently, has not even begun to be addressed (de Lone, 1990).

Thus, while the research demonstration approach has in its first two decades produced a strong record of operational feasibility and credible evidence, and an attentive, influential audience, issues of concern remain--primarily the continuing search for long-term impacts from demonstrated interventions and the replication of those interventions that do have success. Those concerns merit study and discussion. One way to begin is to examine those issues in the context of an actual research demonstration.

The Summer Training And Education Program (STEP)

The STEP demonstration followed the classic pattern of research demonstrations. It identified dropping out of high school as a major problem resulting in poor labor market performance and low incomes. It aimed to reduce school dropout rates by offering a two-summer employment, academic remediation and life skills program to low-achieving 14- and 15-year-olds from poor families. The program included part-time summer work at the minimum wage; remediation involving innovative curricula and teaching methods, and computer-assisted instruction; and high-engagement summer classes focusing on life issues, such as sexual behavior, drug use, careers and community involvement.

The demonstration required a consortium of 11 funding agencies to support; seven years to complete; and $12 million to develop, install in five locations, and evaluate. In those five demonstration locations, STEP successfully randomized 4,800 14- and 15-year-olds, and implemented a consistently operated program intervention, as evidenced by independent field audits and in-program outcome data.

Because STEP focused so heavily on the summer period, the research paid careful attention to the program's impacts during the summer. These impacts were consistently impressive in reading, math and various life skills. Youth participating in STEP (treatments) had test score increases approximately a half-grade higher than controls in both math and reading, and showed substantial improvement in their knowledge of pregnancy prevention--all generated in a six- to eight-week period during each of two summers.

These impressive summer gains eventually led to the replication of STEP in more than 100 locations around the country. The replication sites produced test scores equivalent to, and sometimes better than, those of the original five demonstration sites. The STEP initiative succeeded in several critical areas where there was little or no record of achievement.

These successes seemed to confirm the original hope that STEP could significantly improve the longer-term schooling, income and parenting outcomes for these youth over what they would otherwise experience. But the long-term research that became available in 1991 showed that several years after leaving the STEP program, treatments were no better off than controls. They dropped out of school at the same rate and showed no improvement in early labor market performance or reduction in rates of teen pregnancies.

This record on long-term impacts, which was disappointing yet typical of a research demonstration, led us to reflect on the STEP experience in terms of the insights and lessons it may hold for social programming in general, and for the practice of research demonstrations in particular. This document offers those reflections, and addresses three broad questions:

1. **Does the STEP experience advance our understanding of what is required to produce long-term impacts on serious social problems like those STEP addressed**? STEP itself did not change lives over the long term, but it did have substantial short-term effects. Do those short-term impacts, and the way STEP was implemented, offer any clues to how we might achieve

long-term improvements in the lives of individuals who already exhibit risk factors, such as poverty and low school achievement?

2. **Does STEP suggest changes in approach to the way we currently carry out research demonstrations, what we expect from them, or how we use them**? Although the mechanics of demonstrations have clearly been shown to be workable, their output, in terms of long-term results, has been consistently weak. Is there any relationship between the way research demonstrations have been built and structured, and weak long-term results?

3. **Does STEP's experience in replication--serving about 20,000 youth in over 100 localities through 1991--provide us with insights about the development and implementation of social interventions**? STEP's large-scale replication is unique in research demonstration history; it took three levels of government, plus the philanthropic and nonprofit sectors, to accomplish its consistent programming and results. What do this uniqueness and complexity tell us about the potential to take to scale any other intervention that shows short or long-term effectiveness?

The STEP experience alone cannot "answer" these basic questions, but it can illuminate the relevant issues, and suggest directions for useful change. This report focuses on STEP to help explore the path of knowledge development and transfer from research demonstrations to public institutions.

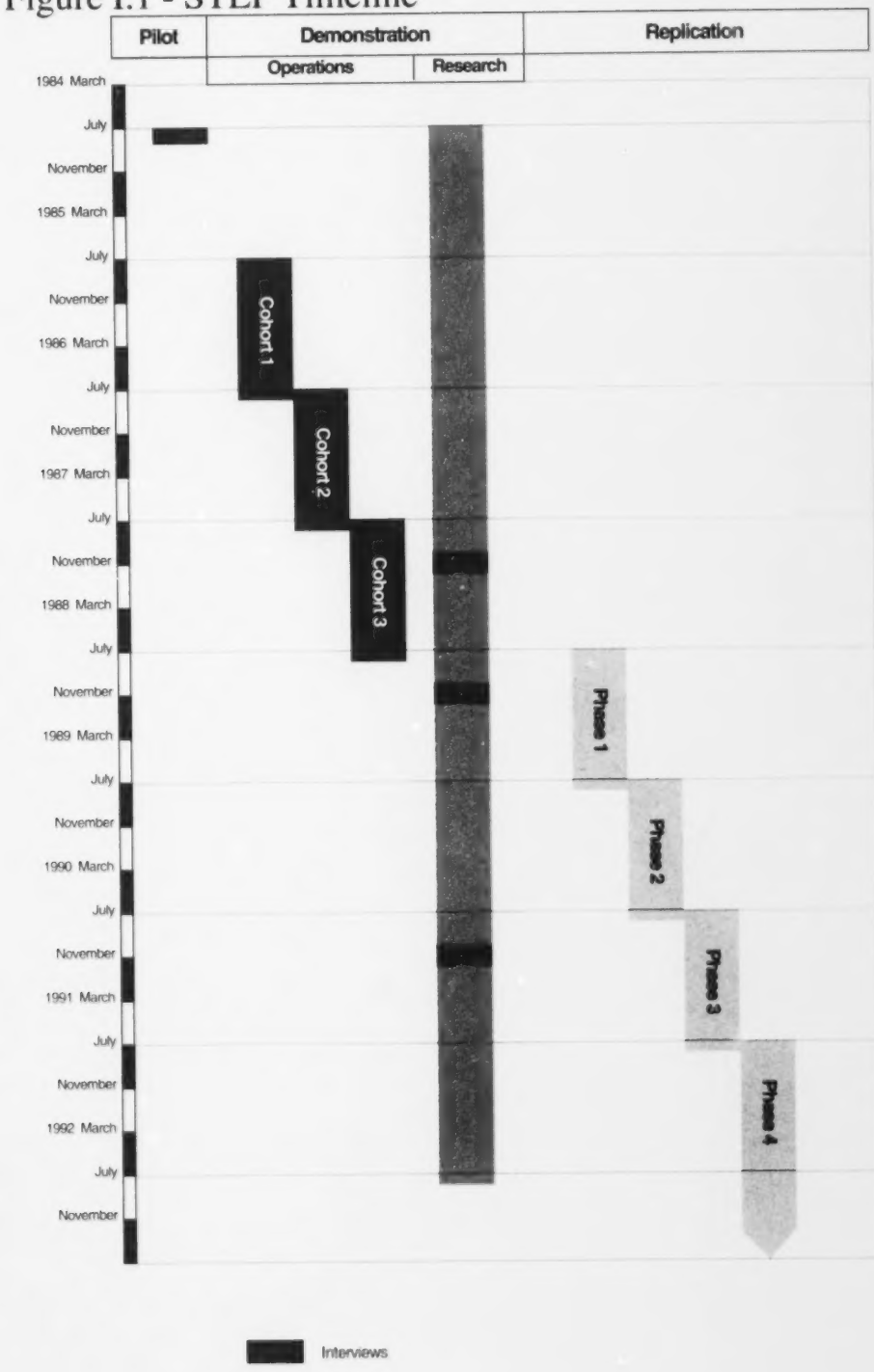
The remainder of this chapter discusses the problems STEP was designed to address and the initial decisions made about the intervention strategy. Chapters II, III and IV chronologically trace the process of transforming STEP from idea to design and implementation and, ultimately, to replication. The program timeline serves to clarify this process. (See Figure I.1.) Chapter V focuses on the long-term research findings[1], and the final chapter summarizes the lessons in a manner that will, we hope, be useful to policymakers, funders, administrators and practitioners.

The Problem And Its Consequences

Too many of our teenagers are failing in school, having children too early and becoming disconnected from society. The numbers are staggering. According to Department of Education statistics--the most conservative estimates--each year over 400,000 youngsters between the ages of 15 and 24 leave high school. In 1989, more than four million--12.6 percent of the nation's 16- to 24-year-olds were dropouts (National Center for Education Statistics [NCES], 1990).[2]

1. Detailed information on the research methodology and the outcomes of specific measures is available in the STEP research reports (1985-1991). Detailed operational information can be found in the STEP Operations Manual.
2. Statistics on the numbers of high school dropouts range widely. Most reports cite dropout rates of 20 percent or more for different subgroups. Advocates contend that real dropout rates for minorities in large urban areas reach 70 to 80 percent. The popular press often uses the figure of one million dropouts a year. These higher rates represent the proportion of students who do not graduate in their 12th year of schooling. Since some of these "dropouts" (those who were retained in grade) complete school in the 13th or 14th year, these higher rates tend to overstate the proportion of students who, in a year, leave school and never receive a diploma. The "event" rates reported by NCES represent the proportion of students who drop out of school in a particular year.

Figure I.1 - STEP Timeline



Statistics tell an even grimmer story for poor minority youngsters in urban areas. According to NCES, suburban and nonmetropolitan white students aged 15 to 19 comprise the bulk of dropouts, but Hispanics, blacks and older students living in major urban areas are more likely than all other students to drop out during the last three grades of high school. The average annual dropout rates for Hispanics and blacks are 7.9 percent and 6.8 percent respectively, much higher than the national average of 4.5 percent.

Other research has also identified significant gender differences among dropouts in urban schools: Hispanic females have a higher dropout rate than Hispanic males, while black males have a much higher dropout rate than their female counterparts (Weis et al., 1989; Natriello, 1987).

Research confirms the importance of a high school education in escaping poverty. In the early 1980s, the average high school dropout could expect lifetime earnings that were only two-thirds those of the average high school graduate (U.S. Department of Commerce, 1983). The substantial gaps between high school graduates and dropouts in earnings and employment are growing steadily. Between the early 1960s and the early 1980s, the disparity in the two groups' earnings nearly doubled among 18- to 24-yearold males (Berlin and Sum, 1988). During roughly the same period, the percentage of dropouts who were employed declined significantly, from 53 percent to 40 percent. There is no evidence these trends have changed in the '90s; most experts believe the gap continues to grow.

Increasing the educational attainment of youth at risk of dropping out of school should improve their employment opportunities and their ability to be economically self-sufficient. Numerous studies have documented that additional years or months of schooling result in additional income (e.g., Griliches and Mason, 1972; Mincer, 1974; Willis, 1986). Whether high school graduation itself adds to an individual's earnings, apart from the additional year of education it embodies, is still unclear. (See contrasting views of Hungerfrong and Solon, 1987, and Card and Krueger, 1992.) However, keeping youth in school longer is unquestionably beneficial (Angrist and Krueger, 1990).

Causes And Associations

The reasons such large numbers of young people drop out despite compelling economic reasons to remain in school are varied and interrelated. Some arise directly from their life circumstances: the NCES survey cited earlier identifies two indicators that are broadly supported by other research--low family income and having dropout siblings.

Weis et al. (1989) and Natriello (1987) argue that youth of lower socioeconomic status have a greater tendency to drop out than their counterparts from privileged families because they are raised in home environments not conducive to learning. Natriello lists minimal educational support, fewer study aids, little opportunity for non-school-related learning, and lack of parental interest in children's progress as indications that learning essentials are missing. Weis et al. also underscore risk factors from family life with the observation that: "Quitting school often caps a series of unfortunate experiences....Nonacademic problems usually precede the decision to leave."

Other predictors of dropping out are more proximate. Young men most often say they dropped out because of negative school experiences, behavior problems or a job. Young women most often cite pregnancy or marriage (Natriello, 1987).

Doing poorly in school is a strong predictor of dropping out (Institute for Educational Leadership, 1986). Approximately 42 percent of dropouts surveyed by the High School and Beyond study report receiving primarily Ds in their classes. Students who are behind their modal (age-appropriate) grade are also less likely to complete high school (NCES, 1990).

In addition, poor youth experience long, empty summers when the limited learning of the school year comes to a halt. "Summer learning decay" is a fact of life for urban teachers, who speak of September as the "catch-up" month. The work of Barbara Heyns and others has produced credible evidence documenting the extent to which disadvantaged youth lose more ground academically over the long summer vacation than do other students. Some have suggested that as much as 80 percent of the difference in school performance between advantaged and disadvantaged students may be due to summer learning loss (Pelavin and David, 1977; David and Pelavin, 1978; Hammond and Frechtling, 1979; Heyns, 1987). Regardless of how we apportion the difference, their findings clearly show that summer learning loss is a critical factor in education.

Many young people who drop out of high school say they do so to go to work; in addition, most would prefer to--and, if they are poor, often need to--earn money in the summer rather than attend remedial classes. While work experience during high school pays off in future earnings and employment (Meyer and Wise, 1980), the effects are strongest for youth who graduate from high school (Berlin, 1983; Berlin and Sum, 1988; Taggart, 1981). Furthermore, results of the Youth Incentive Entitlement Pilot Projects demonstration indicate that providing dropout-prone youth with jobs during the school year and summer does not help keep them in school (Gueron, 1984).

Adolescent pregnancy, a social problem with tremendous economic and personal costs, also has critical impacts on schooling. Pregnancy is the reason girls cite most often for dropping out: only half of all adolescent mothers finish high school (Coombs and Cooley, 1968; Presser, 1978; Hayes, 1987). Girls who become pregnant while in junior high school or high school "complete on average fewer years of school, are less likely to earn a high school diploma, and are less likely to go on to college...than those who delay childbearing until their 20s" (Hayes, 1987). Teenage mothers are more likely to become single heads of household; they are over 50 percent more likely to receive welfare. And while the consequences of adolescent parenthood weigh most heavily on girls, fathers are also affected. Thirty percent of young men who father children at an early age never finish high school (Hayes, 1987).

The causal relationship between teen pregnancy and educational attainment is unclear. While many young women drop out of school because they are pregnant, there is evidence that many adolescent mothers drop out prior to pregnancy and that becoming pregnant may be preceded by poor academic performance (Hayes, 1987). As Dryfoos (1985) points out, "If motherhood follows unsuccessful school experiences, it may be possible that upgrading the quality of education would have an impact on fertility rates."

A number of studies attest to the difficulty of a successful intervention once

the first pregnancy has occurred. A series of evaluations of Project Redirection (Quint and Riccio, 1985) led the researchers to conclude that in light of the complexity of the problems faced by young people once they become parents, strong support should be given to program efforts aimed at preventing the first pregnancy.

In short, the causes of dropping out are both varied and complex--composed of "deep" factors relating to family history, family income, neighborhood and school characteristics, and more "proximate" factors relating to youthful behavior and choices. The research data and analysis available, both when STEP was designed and now--eight years later--provide little or no direction regarding these questions: which or how many of these factors must be addressed to change a youth's decision to drop out; at what age might an intervention addressing one or a group of these factors be most effective; or how long an intervention is needed to produce effects.

The Intervention Strategy

In designing an intervention for research demonstration testing, the initial task is to select a factor, pattern of factors, or group of factors associated with large numbers of people affected by the problem (in STEP's case, dropping out). As noted earlier, the selection process in STEP was aided by available research; however, the data provided no obvious choices. The data did indicate that there was likely to be no single simple solution; common sense added that the combination of deep and proximate factors made it unlikely that any intervention that did not address immediate needs for educational skills, information about long-term consequences of school and social behavior, and income would work.

The selection of an intervention design is influenced not only by the information and research available on causes and associations, but also by what seems operationally feasible and what financial resources are available. STEP was designed in the early 1980s, when funds for social programs were generally being reduced and there was little political and public interest in creating large-scale or resource-intensive interventions for poor youth. Thus, the context for defining what to address in STEP, and how to address it, was one of limitations.

These contextual limitations pushed toward designing an intervention model that was relatively inexpensive, and that could build on existing institutional capacities and funding streams. STEP's design was also dictated by the research demonstration approach, which is based on a scientific, apolitical model of problem-solving: its techniques of random assignment, denial of intervention services to controls and use of statistical techniques to compute results, work best when the intervention strategy is relatively tightly defined.

Thus, available research information, political and financial limitations and the research demonstration approach itself all worked toward selecting proximate, not deep causes of dropping out. Common sense plus available research information worked toward selecting a group of those factors to treat, rather than a single factor, and toward treating those factors over an extended period of time. STEP's creators thus selected two factors for special intervention: the summer period, when learning loss for poor youngsters is high, and early parenting, which is not only clearly associated with dropping out of school and poor adult labor market prospects for both young mothers and fathers, but also seems likely to perpetuate poor prospects for their children.

Choosing the summer period for special intervention allowed STEP's creators to build on a nationwide institutional vehicle, the Summer Youth Employment and Training Program (SYETP), for carrying out both the research demonstration and any later replication initiative. The summer jobs program had been in existence since 1965 and provided a paid work experience for between 600,000 and one million poor youngsters each year. Funding for the summer jobs program, administered through the U.S. Department of Labor, had survived the various major legislative changes that governed the nation's employment and training assistance to disadvantaged citizens, most recently in the 1981 shift from the Comprehensive Employment and Training Act (CETA) to the Job Training and Partnership Act (JTPA). Thus, the federal summer jobs program offered a stable and experienced institutional base on which to operate STEP.

In addition, the summer jobs program offered a substantive benefit: it provided paid work experience for poor young people, whose employment and income opportunities are limited. There is broad policy consensus that summer work experience is good for young teenagers. It begins their connection to the labor market's practices and expectations--in short, it helps produce better adult workers. Although there is inconclusive research evidence on summer work's long-term effects, its common sense and broad appeal, combined with poor youth's need for summer money, made it an important third element in the STEP substantive strategy.

The federal summer jobs program had been consistently criticized for being no more than a way to keep poor youth off the streets and put a little money in their pockets; it did not attempt to build educational skills, and many critics felt it did not offer a solid work experience. When STEP was being designed, pressure was already building for new legislative requirements that would make the summer jobs program a more positive experience for youth. Thus, STEP's development dovetailed with a broader movement for change and held out the prospect of influencing and improving an existing national system of youth services.

STEP's designers did not build on another national institutional vehicle that was obviously interested in finding a solution to the dropout problem: the public schools. The professional experience of STEP designers was with employment and training systems, not the schools; that experience and their own interests dictated use of the employment and training system as the entry point for STEP's creation and demonstration. While it was clear that STEP's operation would require school cooperation, gaining this would ultimately be left to the local employment and training agencies. Adding to the difficulty of securing school involvement in STEP was the federal Department of Education's (DOE) disinterest in new initiatives in the mid-1980s, when that Department was concerned about its own survival. While the Department of Labor (DOL) supported STEP from early on and made it a priority in its research and development agenda, STEP's developers were unable to gain a response from DOE officials regarding possible joint sponsorship of the demonstration.

The decision to choose pregnancy prevention as the second major focus of STEP meant that the demonstration could benefit from working on a high-profile issue in serious need of innovative programming. Although there were numerous small pregnancy-prevention efforts around the country, few had produced solid results, and most were bootstrap operations with little investment in curriculum development and training for program implementers. It was an

area ripe for systematic development and testing.

In addition to these decisions about the demonstration's substantive components and their delivery system, STEP's designers also had to define who would be eligible for the intervention. Findings from previous demonstrations, in addition to the judgments of many policy experts, administrators and researchers, indicated that program interventions needed to focus more on early prevention.

However, the use of the employment and training delivery system set the earliest age for STEP participants at 14, the legislatively set threshold age for participation in the summer jobs program. Data indicated that few local delivery systems were in fact choosing to serve youth that young; summer jobs programs were reserving most of their jobs for older youth who, administrators felt, needed the work and income more. Thus, STEP was aimed exclusively at 14- and 15-year-olds whose families met the SYETP poverty guidelines. In addition, to be eligible, a youth had to test at least one grade level behind in either math or reading.

Thus, although STEP was preventive compared to previous demonstration models, which had aimed mostly at older youth and adults, it was in fact barely so. The eligible population was very near the age when large numbers of poor youth begin to drop out and become young parents, when they are already doing poorly in school, and are already at the age of sexual development and exploration.

STEP's creators then, influenced by various limitations, picked two of the many possible substantive areas--summer learning loss and teen pregnancy--for intervention to improve school performance and prevent school dropout. For practical reasons, they engaged the federal employment and training system to carry out the demonstration, aiming at its youngest eligibles. These choices made it possible to garner the necessary financial support, to design and mount the model within a brief period of time, and to operate the model (and later to replicate it) with consistency and efficiency. The practicality of the choices neither improves nor undercuts their conceptual soundness. In fact, STEP's creators were following in a well-established tradition of scientific experimentation: using materials available to the particular experimenters to make an informed guess about what might work.

II. Program and Research Designs

A useful test of an intervention strategy requires a soundly developed program model and a rigorous research design. The previous chapter discussed the limitations of knowledge, interest, institutional capacity and financial resources that shaped the intervention strategy for STEP. Those same limitations would affect the basic structure of the program model design, but would not significantly affect the resources devoted to the development of innovations for the various STEP program components, or to its research design or implementation.

Program Model Design

STEP's creators had decided to focus on stopping summer learning loss and preventing teen pregnancy, and to integrate those activities into the structure of the federal summer jobs program. The challenge was to design a program that could work within the summer jobs structure and still offer the prospect of having a significant impact on youth.

It seemed likely that one summer was not sufficient to have an enduring effect on participants. This judgement was based less on evidence than on experience in working with at-risk youth, combined with common sense about what it takes for humans to sustain serious changes in their life course.

However, the JTPA system, which operates the summer jobs program, had very little experience in linking its summer work program to any larger or longer intervention. In addition, the long experience of the summer program has established a typical cost of $1,000 per youth.

STEP designers thus anticipated some reluctance or resistance from the JTPA

system in expanding the summer program. To ensure that enough local operators would be interested in testing the STEP idea, the designers decided to stretch the system substantially, but not to push too far beyond the system's *modus operandi* and financial limitations. It was decided that STEP would operate for two summers; summer jobs program administrators would thus have to commit that some youth could return and get a second publicly supported summer job. They would also have to agree to operate a program that, with its educational remediation and life skills components, would cost 60 percent more per youth than the usual cost each summer--$1,600 versus $1,000. In short, they would have to agree to spend more than $3,000 for one youth, where before that amount might have provided single summer jobs to three youth. They would also have to agree to a program whose operation was considerably more complicated than a summer jobs program.

As shown in Figure II.1, the STEP model in its final design consisted of five days a week of work and classes for six-to-eight weeks in two consecutive summers. Each summer's program included approximately 200 hours of program involvement, broken down into:

- Work--90 hours: STEP jobs are half-time, but otherwise similar to those usually available in the summer jobs program. The jobs may be group projects or individual placements: for example, aides in day camps, day care centers or parks and recreation programs, buildings and grounds maintenance, or office work.

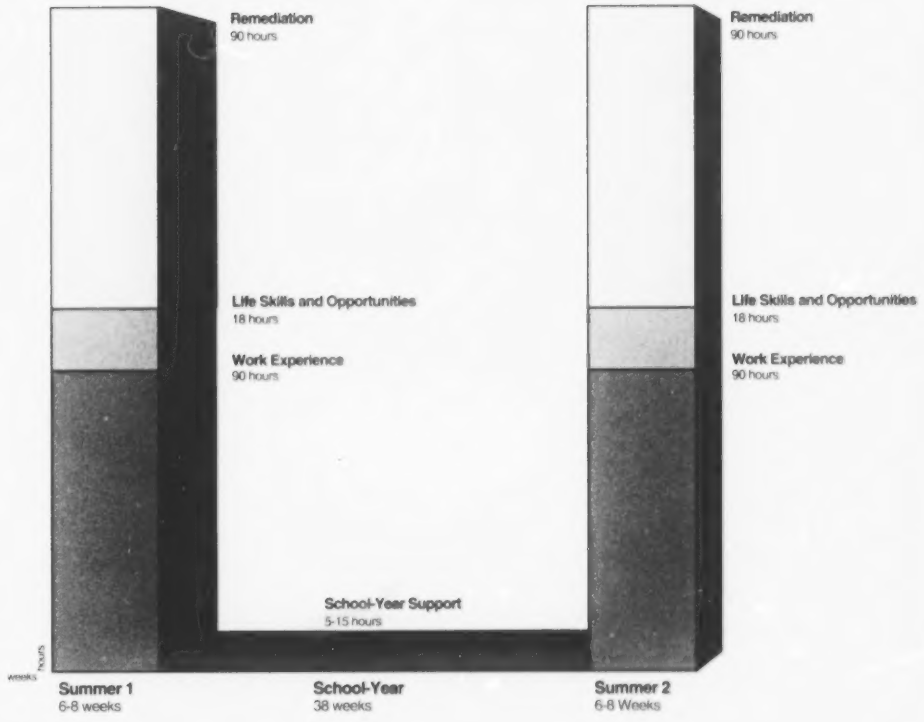
- Remediation--90 hours: STEP youth spend every morning during the program's two summers learning reading and math skills and higher-order thinking skills, such as problem-solving. (By the second year of the demonstration, the remediation component had evolved into the *Practical Academics* curriculum, which combines specially tailored teaching modules, computer-assisted instruction, practical exercises, journal writing and sustained silent reading).

- *Life Skills and Opportunities (LSO)* classes--18 hours: Two mornings a week, STEP youth attend *LSO* classes, where they learn about decision-making and the importance of responsible social and sexual behavior, avoiding substance abuse and maintaining good health. The goal is to give them more confidence in their own capacity to deal responsibly with the future.

STEP youth were to be paid for the time they spent in the classroom as well as for their part-time jobs. The designers felt that poor youth would not voluntarily choose to enroll for part-time income if full-time income was also available. The JTPA system is a voluntary system, and there was no mechanism or precedent for denying its regular services to a particular group of eligibles. In addition, the research designers indicated that it would be important to limit the differences between controls and treatments to those program elements aimed at changing behavior. Thus, STEP participants, both controls and treatments, received the same wages over the summer.

The intervening school year itself was not addressed by STEP's designers. The model design, however, included a limited school-year component (school-year support) whose main goal was to get youth to return to STEP's second summer. A youth's participation in school-year support activities was voluntary. Within guidelines provided by the designers, most STEP demonstration sites put into place a very modest school-year component, composed of

Figure II.1 - STEP Program Model



limited one-on-one adult contact, recreation and other non-educational activities. Most youth spent between five and 15 hours in those school-year activities.

STEP's involvement in a youth's life is quantitatively modest relative to the total time spent interacting with other major influences, such as family, school, peers and neighborhood. The remediation work in one STEP summer, for example, amounts to less than 10 percent of the time spent in a regular year's school work. Thus, STEP is best seen as an inoculation, a booster shot intended to initiate a process within each youngster that will continue and result, over time, in the learning, work habits and life choices that lead to good health and self-sufficiency. The learning component is intended to prevent summer learning loss and, thus, serve as a booster for the regular school year.

The STEP model is also qualitatively modest. It makes no attempt to influence or interact significantly with the regular schoolyear structure, content or teaching methods, or with the youth's parents, friends, peers, neighborhood or neighborhood institutions. STEP deals primarily with its participants' mental abilities, especially their knowledge, curiosity and reasoning. It does not attempt to change any aspect of their environment. STEP's qualitative modesty defines it as a self-improvement intervention that takes as an unchanging "given" the dominant influences in a youth's life.

Given that these other, non-STEP aspects of life significantly influence a youth's behavior over time--and that the force of those influences in many poor urban neighborhoods does not work to keep youth in school--STEP is an optimistic, as well as modest, initiative. As noted earlier, the optimism and modesty were shaped by various contextual limitations, but also represented a reasonable, cautious approach to the process of finding solutions. Since little was known about what would reduce dropping out, it could be argued that the sensible first step in the testing process was to see if a relatively limited, low-cost intervention would produce positive results.

But while the modesty of STEP's "booster" approach is clear when compared to the total time and number of influences that exist in a youth's life, it is not modest compared to most special interventions available to poor young teenagers. Summer "Fresh Air" camps, day trips and other such activities are available in some poor communities for some youth; almost without exception, however, these do not involve either the concentrated amount of time in a variety of activities, or the focus on education, work and life skills that STEP has. Boys and Girls Clubs, Police Athletic Leagues, churches and other voluntary community youth services have greater potential than STEP for engaging a youth over a long period of time, but few of these offer paid work experience or a curriculum of learning and life skills tailored specially for young teenagers. In short, STEP's services are not widely available, either singly or in combination, in poor communities (Littell and Wyna, 1989).

Thus, relative to other available interventions and supports, STEP is not so modest. In fact, in many of the communities where STEP has operated, it is seen as a major initiative for youth at risk of dropping out. STEP may be a modest "booster shot" in a youth's life, but it is a significant expression of public commitment to assisting poor, young adolescents.

Investment In Innovations

Prior to STEP, most social research demonstrations consisted of program models that did not require substantive innovations. Social demonstrations had previously focused on work experience, preemployment skills, and training programs that sometimes added conventional support services like counseling, day care and transportation assistance. Innovation occurred primarily in the structure and coordination of existing activities. New program components were left to local operators to develop.

STEP, however, attempted to combine a group of substantive activities, most of which were innovations, within one coherent program model. Thus, even though STEP was a modest intervention in its participants' lives, it was not modest in innovation. It required what, for the social intervention world in the 1980s, was a significant infusion of resources in innovation development.

The *Practical Academics* and *Life Skills and Opportunities* curricula--STEP's summer learning components--are its major innovations. The simplest design would have been to require eligible youth who wanted summer jobs to attend existing summer school programs. But advice from education experts and the available data indicated that regular summer school programs for this age group were generally limited to specific course offerings that did not aim to improve basic skills. There was also an emerging body of educational evidence that students having difficulty in school benefit from curricular materials that have relevance to their daily lives, and from teaching methods that encourage student engagement and participation, neither a regular practice of public summer schools. Thus, the quality and availability of existing institutional capacity became major considerations in designing the education component.

Given these facts, the decision was made to design a new approach to summer remediation. Originally, the idea was to provide guidelines and assistance to the demonstration site operators, who would flesh out the framework, develop their own curriculum, locate materials and implement the program. The designers intended that local employment and training agencies would arrange with local schools to carry out this process, as well as provide space for the summer education program. The designers decided not to specify the curriculum and actual materials in detail since, they believed, local teachers and school districts would most likely resent such prescription. In addition, pressures of time--the demonstration had to get started on a schedule acceptable to its major funders--did not allow for development and testing of a detailed curriculum.

At the same time, STEP's designers did decide to create a full curriculum for the pregnancy-prevention component. While some related materials were available, they had significant drawbacks: none was tailored to STEP youth's specific needs or covered all the various life issues that were felt to be critical for disadvantaged, mostly minority 14- and 15-year-olds, nor did any deal with emerging issues like AIDS. In addition, none was compatible with the structure of the summer jobs program. Since the issues to be covered were so sensitive, there was also a clear need for a well-documented, professionally administered training program for the people who would be teaching this component. Thus, the new curriculum, called *Life Skills and Opportunities*, and its accompanying training materials were fully developed at the central level.

Developing these two components required a major commitment of resources relative to the usual public and philanthropic investments for interventions aimed at well-defined disadvantaged groups. This need for investment increased after each of the first two summers of STEP as it became clear that the sites wanted, and would benefit from, a full, innovative remediation curriculum, not just the framework and guidance that had been offered.

Since there were few individuals or groups that had done work in these areas, STEP's innovations required outreach to many organizations and experts. P/PV's staff did some of the conceptual and design work on the *Practical Academics* curriculum, but *LSO* relied almost solely on outside expertise. This reliance on multiple consultants and organizations suggests the level and variety of resources necessary to create substantive innovations, resources that are not present together in any one organization in the social intervention field, or in any one community.

The financial investment in STEP curricula was sizeable from the perspective of the amount of time these innovations would directly involve themselves in a youth's life. Over the first two years of the demonstration, approximately $500,000 was spent developing the remediation and life skills learning packages. Yet the remediation component was to account each summer for only 90 hours (or two-and-a-quarter regular work weeks) of participants' time, and the *LSO* class for only 18 hours.

While STEP's development investments were substantial, however, they were aimed at those elements that represented the intervention's major targets among causes of, or antecedents to, dropping out: summer learning loss and teen pregnancy. There was very limited financial investment in developing new approaches to other components of the demonstration, including summer work experience. Since local operators had experience with developing and operating summer jobs programs (though they had little experience in hiring 14- and 15-year-olds or in developing part-time summer jobs), STEP's creators felt that national guidance or development in these areas would add little to what local efforts could accomplish. Similarly, the school-year support component was left to local development, though less because of perceived local experience and development skills than because it was not viewed as a major program component. STEP youth were not required to participate in STEP school-year support activities, which were intended as a connector between the two STEP summers--a way to keep participants interested in coming back for the second summer.

In sum, because STEP was a major research demonstration and centrally organized, it was able to attract substantial funds for investment in substantive innovations. The attractiveness of these innovations to investors, both public and private, was that they were aimed at long-term changes in youth's lives.

Research Design

While the very word "research" may suggest to some an arcane activity with little direct relationship to common sense and practical questions, the STEP research design outlined in Figure II.2 is a model of simplicity and usefulness. It aims to answer two basic questions:

1. Does the program stop summer learning loss and generate increases in conventional test scores in reading and math and school performance, and in knowledge and behavior related to pregnancy

Figure II.2 - Summer Training And Education Program Research Design

GENERAL STRATEGY

- To assess the short-term and long-term impacts of the program on participants.
- To assess the feasibility of implementing the model in various settings and on a large scale.

EXPERIMENTAL DESIGN

- Random assignment to treatment and control groups.
- Control group participates in the local site ***SYETP*** program. At two sites (San Diego and Seattle) control youth are also guaranteed an ***SYETP*** job for the second summer.

SAMPLE SIZE

- 4,800 14- and 15-year-old economically and educationally disadvantaged young people
  (2400 treatment/2400 control youth).

COMPONENTS

- ***Implementation Analysis (1984-1988)***: Examines the processes involved in planning, coordinating and operating the program. Uses qualitative case studies and structured observations by P/PV staff and life skills and remediation experts. Quantitative data were obtained from application forms, questionnaires and program records.
- ***Impact Analysis (1985-1993)***: Employs an experimental design involving random assignment. Examines the effect of the program on schooling and academic performance, adolescent pregnancy and parenthood, and early labor market experience.
  - *In-program (1985-1988)*: Examines results of summer tests (the Metropolitan Achievement Test) administered to all treatments and controls the first summer, and to all treatments (and to controls at the two sites that guaranteed them a second-summer job); and participant questionnaires administered simultaneously with the MAT, as well as a range of program records.
  - *Long-term (1987-1993)*: Examines data from school records annually and from three waves of participant interviews.

prevention, good health practices and other life skills?

2. If so, do those increases in performance and knowledge, and changes in behavior, endure over time, years after the program, and result in reductions in rates of dropping out and teenage pregnancy, and in other positive impacts that are predictive of a successful adult life in American society?

To answer those questions with scientific reliability required development of an experimental design involving random assignment of eligible youth to treatment (STEP) and control (SYETP) groups. Youth in the control group would have jobs for at least one summer (two sites provided second-summer jobs to controls), but would not be receiving the STEP intervention. STEP's creators and funders believed random assignment was critical to the research strategy because it would yield statistically equivalent groups at program start, allowing the researchers confidently to attribute different outcomes between groups to STEP, rather than to other outside factors. Thus, the impact research compares STEP to SYETP and does not speak to its value relative to an empty summer experience, the status quo for most disadvantaged young teens.

Assessment of the effect on participants required multiple tasks: the development of baseline and postprogram questionnaires, testing procedures, randomization procedures, confidentiality procedures, and analysis equations. The data sought from the youth had to be sufficient to allow answers not only to the basic questions but also to derivative questions. (For example, did gender or race have any relationship to performance?) The need to measure changes in knowledge and behavior related to sexuality meant that new research instruments had to be developed. Research on STEP's school-related and dropout-prevention goals required extensive data collection from participants' school records and several waves of personal interviews--and a significant follow-up period after youth had completed the program. In addition, the effects of STEP on individuals would be given a context by an implementation study that described the process of putting the program in place and determined the possibility of large-scale replication. The research apparatus cost about $400,000 to develop and almost $7 million over eight years to carry out and complete.

Although developing the research design for STEP was a formidable task, it was more precise and more guided by known rules than was the process of developing the demonstration program model. This precision in knowing what it took to get scientifically credible results helped ensure that the financial limitations that shaped the intervention strategy and substantive model decisions for STEP had little or no effect on the resources put into the development of the research model or in carrying it out. The common interest of funders and policymakers was in getting highly credible results. In addition, the research costs were one-time, and thus had no implications for program operating costs, policy trends or long-term public budget considerations.

Thus, STEP's design effort can be characterized as intensive in innovative depth and research thoroughness, but more cautious and influenced by external restraints in the extent of the intervention's involvement with its youth. This imbalance comes as no surprise to anyone interested in social policy: the basic program intervention, model shape and costs must, even in a test mode, fit within the larger context of institutional capacities and available funds.

If large-scale replication is to be possible, there are additional implications regarding feasible cost and program structure. The functions carried out by P/PV as the research demonstration manager--design, research, program installation and maintenance--involve one-time costs that have typically been borne in demonstrations by a consortium of funders, and rarely by the public system alone. They are, to put it simply, easier to fund adequately and to do properly.

Although hindsight always improves insight, it is possible to see in STEP's early decisions--about the intervention strategy, program and research designs, and innovation investments--the seeds of its achievements: consistent results across sites; an appealing and easily-installed program, well-designed for broad and consistent replication; and consistently positive immediate results. They also bear the seeds of its major disappointment: the general lack of positive long-term impacts.

III. Testing The Model

STEP was transformed from initial idea into two pilot programs in just over four months--March to June 1984--and into a five-site national demonstration one year later. This swift pace of development was accomplished by a highly compressed process of planning, materials development, site selection and training.

Such a schedule would never have been set, or met, had it not been for STEP's focus on the summer period. The program had a brief operational window once a year; missing the 1984 summer would have entailed a 12-month delay. In addition, the designers' early decision to provide a set of remediation principles and guidelines rather than a full curriculum eased P/PV's early development process.

The Pilot

The pilot period was an important test of STEP's overall operational viability. There were no solid operating precedents for a highly structured, multicomponent summer intervention for such a young group of at-risk students. In addition to this general issue, there were very specific concerns. Would the number of 14- and 15-year-olds opting to enroll in STEP be sufficient to justify carrying out a complex research design that included random assignment and years of follow-up? Could the "educationally needy" eligibility requirement be successfully and uniformly operationalized? Could the academic tests be carried out with enough uniformity between controls and treatments--and would enough controls take the tests--to produce credible test comparisons? Would the remediation curriculum be carried out at the local sites with enough uniformity and effectiveness to justify a major research demonstration? Would the *LSO* curriculum, which contained discussion of sensitive sexuality and pregnancy-

prevention issues, be acceptable to parents, schools and local communities?

Two local employment and training agencies--in Boston, Massachusetts, and Pinellas County, Florida--were selected to operate the pilot. The tight timeline precluded a competitive selection process; sites were selected based on P/PV's judgment of their interest and capabilities in working to develop the program locally and their ability to commit JTPA funds in time to implement STEP that summer.

The fast pace helped promote an early division of development functions between P/PV and the local sites. P/PV focused its early design work on the *LSO* curriculum, academic tests and testing procedures, and the research design and early procedures and instruments necessary to ensure its success. The two sites focused on developing part-time work for 14- and 15-year-olds, hiring teachers for the remediation component who would follow the STEP guidelines and teaching principles, and finding teachers who would be comfortable and competent addressing *LSO* issues. All these were new activities for local employment and training operators.

The 1984 summer pilot had mixed operational results. Each site was able to attract close to its goal of 500 students (250 treatment/250 control), but both felt that doing so had required extraordinary recruitment efforts. They also felt that the number of job slots allocated to 14- and 15-year-olds was too large and had forced them to shift their focus away from the older youth traditionally served. They believed that smaller numbers of students would produce a higher quality program more quickly--an important consideration for a research demonstration, which allows very little operational seasoning time before beginning the research.

The remediation component proved difficult to implement with any degree of uniformity. Without a common curriculum or tightly specified content, remediation locations followed different curricula and used widely different instructional techniques. However, the *LSO* curriculum, despite its newness and sensitive nature, was well-implemented and was received enthusiastically by students and teachers.

The part-time work experience was also well-developed and implemented. The employment and training agencies were experienced in worksite development and administration; they adjusted quickly to the needs of younger students, the challenges of developing part-time worksites, and the logistics of getting students from classroom to worksite.

The pre- and posttests of reading and math skills were successfully implemented locally, with good participation by youth in the control groups, though there was no way to determine whether controls had scored poorly because they were not motivated to do well on the test. The results were sufficiently impressive--between 0.5 and one full grade equivalent difference between the groups in both reading and math skills--that P/PV and The Ford Foundation, which funded the pilot, determined that it was worthwhile to proceed to a full demonstration.

The pilot's mixed operating record did call for adjustments to the model. While the *LSO* curriculum was left substantively intact, the remediation component was to be centrally developed into a more tightly specified curriculum process called Community Organized Remedial Education (CORE). It was clear that greater central assistance and direction were required to yield a substantial degree of uniformity across sites, but CORE would still allow local

responsibility and discretion to ensure that local teachers felt ownership of the curriculum. In addition, a financial incentive plan was developed for the pre- and posttests to ensure that controls and treatments would take the test with equal motivation. The plan was a lottery, in which winners would get $50 in cash; the number of tickets students had in the lottery depended on attendance at the test and the number of correct answers.

The Demonstration

Research planning indicated that a sample of 3,000--1,500 controls and 1,500 treatments--would provide sufficient data to address with confidence both the basic research questions and their derivative issues. It was decided to reach that number by engaging five local operators to enroll 330 students (165 treatments, 165 controls) in each of two consecutive summers. Since STEP is a two-summer program, each site would contribute 660 students--in two overlapping waves, or cohorts--to the overall sample. This number would be sufficient to support a site-level analysis as well as sex, age and race analyses of the pooled sample.

STEP participants and controls would be tested before and after their first summer in the program. STEP participants and controls at the two sites that promised second-summer jobs to both groups would also be tested at the beginning and end of the second summer. After the treatments had completed two STEP summers, the full sample (treatments and controls) would be interviewed at several points for up to four-and-a-half years (six months after their expected graduation). Throughout, data for both treatments and controls would be collected from school records.

Site Selection

A national competition was conducted, and 15 of the 595 local employment and training agencies who then administered SYETP applied to operate STEP. In March 1985, five sites were selected to implement the STEP demonstration: Boston, Massachusetts; Fresno and San Diego, California; Portland, Oregon; and Seattle, Washington.[3] (See Figure III.1.) At the time of selection, they had three months to prepare for the operation of STEP.

It is significant that so few local employment and training agencies--less than three percent of eligible operators--applied to participate in the STEP research demonstration. One drawback was the financial disincentive. Each site was offered $25,000 to assist in covering incremental operational costs beyond those associated with the regular summer jobs program; however, those incremental costs were estimated to be approximately $200,000 per site for each of the two cohorts.

But despite this drawback and some reasonable reluctance to participate in a test that might not work, sites were being offered a new approach to summer programming at a time when the summer jobs program was subject to serious ongoing criticism. In addition, legislative efforts were already underway to amend the summer jobs program to require educational remediation as an integral part of its program. (This requirement became law in 1986.) Thus, STEP represented an at least partially funded head start on a programmatic requirement that was gaining support and seemed to make sense.

3. A sixth site, Broward County, Florida, was selected but subsequently did not take part in the demonstration because of education/employment and training coordination difficulties vis-a-vis the LSO-related research.

Figure III.1 - Summer Training And Education Program Demonstration Sites' Operating Agencies

Service Providers by Program Component	**Boston, MA**	**Fresno, CA**	**Portland, OR**	**San Diego, CA**	**Seattle, WA**
	LEAD AGENCY *Mayor's Office of Jobs and Community Services (MOJCS)*	**LEAD AGENCY** *Fresno Private Industry Council (FPIC)*	**LEAD AGENCY** *Portland Private Industry Council (PIC)*	**LEAD AGENCY** *San Diego Regional Employment & Training Consortium (RETC)*	**LEAD AGENCY** *Seattle Department of Human Resources (SDHR)*
REMEDIATION	Action for Boston Community Development (ABCD) at Boston University, Harvard University, Northeastern University and The University of Mass-achusetts (UMass)-Boston	Fresno Unified School District	Portland PIC and Portland Public Schools	San Diego City Schools	Seattle DHR and Seattle Public Schools
LIFE SKILLS	ABCD at Boston University, Harvard University, North- eastern University and UMass-Boston	Fresno Unified School District	Portland PIC	San Diego City Schools	Seattle DHR
WORK EXPERIENCE	ABCD at Boston University, Harvard University, Northeastern University and UMass-Boston	Fresno County Economic Opportunities Commission	Portland PIC	San Diego Regional Youth Employment Program (REGY)	Seattle DHR
SCHOOL-YEAR SUPPORT	Mayor's Office of Jobs and Community Services	Fresno Unified School District	Portland PIC	San Diego City Schools	Public/Private Ventures in collaboration with Seattle Public Schools

At that time, however, the employment and training system was in conflict over a major issue central to STEP: whether local employment and training operators should be directly supporting educational remediation for in-school students. The majority of operators felt they should not be doing so, holding that remediation, summer or otherwise, was the job of the schools. They were supported in that judgment by their governing boards, which by law were dominated by local business leaders.

The relationship between the employment and training system and the public school system was thus a significant issue for STEP at its inception. (It would continue to be so through replication.) The demonstration guidelines took no position on which institution--the employment and training operator or the local schools--should financially support and/or implement the remediation component. P/PV assumed that the schools and their teachers would participate in developing STEP, and required local school district superintendents' written approval of STEP and cooperation with research information requirements. But it was left to the local employment and training agencies and the schools to negotiate the degree of cooperation and to specify each player's role and level of financial support.

However, the relatively fast start-up schedule depressed any potential STEP might have had to develop significant employment and training/school cooperation at the state or local level. In 1985, few employment and training agencies had solid working relationships with their local school districts. Local commitments to STEP had to be made quickly; thus, there was no time for a measured process of familiarization and negotiation.

The result was that the five chosen sites had relatively one-sided support for their remediation component. In all five, the employment and training system paid the vast majority of development and operating costs, including teacher salaries. The lack of a strong federal Department of Education presence in STEP may have made the local process of gaining public school financial support more difficult; but given the modest federal control over, or financial presence in, local schools, it is not clear that this absence actually had much impact. The schools did contribute to all five sites--primarily in-kind items, such as classroom space and computer use. Nonetheless, we estimate that over 85 percent of all educational remediation costs were borne by the employment and training system.

This pattern softened somewhat in the replication effort, after STEP had proven its value in reducing summer learning losses and when the states were thoroughly involved. But STEP's reputation as an employment and training program has largely persisted; even after the early positive results, the federal Department of Education has remained uninterested in sponsoring STEP as a model summer intervention.

The other notable result of the test-site selection process was the predominance of West Coast cities among both applicants and selected sites. Four of the five cities selected were on the West Coast, even after special efforts were made by P/PV to interest major eastern, midwestern and southern communities.

Although a site-selection process for one national research demonstration does not easily support generalizations about entire regions of the country, it did appear that education/employment and training cooperation was less well-developed in the large, older eastern cities, and that employment and training officials' attitudes about the degree to which their system should support and engage in remedial activities were more

entrenched on the East Coast. With the exception of Boston, whose employment and training system not only worked with the public schools but also with several universities in implementing STEP, the demonstration was a West Coast initiative.

Uniformity and Variation

Common sense, as well as testing methodology, requires that an intervention to be tested in multiple sites must be implemented with a substantial degree of uniformity across those sites. Only then can the research data from each local program be aggregated into overall demonstration findings. Since the reliability and credibility of STEP's research would depend on aggregation, implementation uniformity was critical.

Implementation uniformity is a vital concept, both for the validity of the research and for any later efforts to transplant or replicate model programs beyond the test locations. Uniformity's relevance also extends to the broader, ongoing political debate in America about central versus local control in social programming, and "cookie-cutter" versus local creativity approaches to achieving effective and durable social interventions. But while the need for implementation uniformity in a research demonstration is widely accepted, what it means--the forms it can take--has not been the subject of much analysis.

Besides its status as a test intervention, STEP's compressed operational period (the summer), and its multiple components and innovations conspired to make the program tightly structured. The number of hours per component and the logistical placement of each were prescribed in detail; the five demonstration sites were given no discretion regarding these structural issues.

To some extent, STEP also prescribed uniformity in the content of its components. As noted earlier, the LSO component had carefully defined teaching approaches and materials that did not allow for significant variation. The sensitivity of the issues, in addition to the need for uniformity, had convinced P/PV of the need for such prescription. The CORE process for remediation, while more fully developed for the demonstration than it had been for the pilot, still imposed only a cautious uniformity, leaving considerable discretion to the sites. Only the content of the work experience component, and whatever activities sites wanted to schedule during the school year, were left entirely to local discretion. Overall, the STEP model was, and is, viewed as a highly prescriptive model by those favoring local control.

The demonstration sites generally accepted this prescriptive intervention model and, on occasion, wanted even more definition and concrete guidance. For example, from the sites' perspective, CORE was not sufficient; they complained about the enormous work required to develop a remediation curriculum. The careful development of the *LSO* curriculum and its detailed training regimen were viewed as useful, and more training was sought by the sites.

P/PV responded with more detailed remediation materials for the 1986 summer. The CORE principles were folded into a fuller curriculum called *Practical Academics*. In addition, field observations of the Summer 1985 classes led P/PV to develop guidelines for the preferred backgrounds and skills of the teachers of both *Practical Academics* and *LSO*; and to require a lead teacher in each site, someone who was more experienced in and sympathetic to the approach used in *Practical Academics* and who could both train and oversee teaching performance.

Thus, the STEP model became even more prescriptive during the demonstration, often at the request of the

local operating sites, and always as a result of implementation experience.

This prescription went beyond program logistical structure to program content, implementation techniques, and the structure and personal characteristics of staff. Yet this model-tightening did not produce major disputes between P/PV and local operators.

Why did local operators behave counter to the political paradigm on issues of central prescription and local control? The answer does not lie with financial incentives. As noted earlier, the financial support offered operators was too small a share of the incremental operating costs to have affected sites' willingness to put up with burdensome centralized control. As the model became tighter, STEP grants to sites increased somewhat, enabling them, for example, to hire a lead teacher; but the financial incentives were still limited. And though self-selection into the demonstration played some role in sites' willingness to seek and accept increasing prescription and uniformity, all were chosen because they were strong operators who had records of creativity in programming; they were not a passive group.

Discussions with site operators suggest five primary reasons they accepted a high degree of central control:

1. The lack of local investment resources and capacity to develop substantive innovations on their own;

2. The lack of significant operating experience in the remediation and life skills areas;

3. The back-and-forth review process between P/PV and sites that was utilized to develop central prescriptions and substantive materials;

4. The ability of P/PV to offer local operators not just guidelines and output measures, but professional, hands-on training and technical assistance in those areas; and

5. The balance in STEP between highly prescribed components and others (work experience and school-year support) that were left almost totally to local discretion.

Local operators were not being told how to do things that were best handled by local experience and local knowledge, particularly the part-time summer work experience and school-year support components. And none of the sites had the resources required to develop either the *Practical Academics* or *LSO* approaches or curricula. As in most social program funding allocations, employment and training operators had no separate budget line for innovation development. Their funds went for three basic functions: service operations, ordinary administrative costs, and special administrative costs to meet state and federal reporting requirements. Although these were all outstanding local operators, too much of their available time and resources were consumed to allow them to properly develop the content of an innovation.

In addition, because this was a research demonstration, where presumably as much as possible should be laboratory-perfect, P/PV had raised funds not only to develop innovations in remediation and life skills, but to offer hands-on training, regular observation and feedback to the sites, and specialized assistance when problems arose. In short, unlike the usual governmental central/local relationship in social programs, P/PV was able to go beyond programmatic regulations, administrative requirements and output demands to offer a substantive partnership in carrying out the innovations, and a

collegial, nonbureaucratic style of communication in getting local input for their development.

Over the course of the demonstration, P/PV expended approximately $1 million in developing the *Practical Academics* and *LSO* curricula and implementing them at the five sites. Approximately 50 percent of that was spent on content development, with the other 50 percent spent on implementation assistance and monitoring. This amounts to approximately $100,000 per site for technical assistance and monitoring over a three-year operational period. These numbers provide a rough sense of what is required to install an innovative human service program with consistent implementation and results across multiple sites.

In essence, the special research demonstration nature of the STEP initiative allowed for innovation expenditures that cannot be duplicated at the local level and are not currently achieved at the federal or state levels.[4] These expenditures not only allowed the uniformity requisite for a research demonstration but also produced consistent in-program results, as detailed later in this chapter.

In addition, the innovation resources available to STEP allowed for a central/local relationship that transcended the usual political debates over central versus local control, and "cookie-cutter" versus local creativity approaches. The STEP operating experience indicates that there is always room for creativity and discretion in the direct delivery of services to young people. It is simply not possible to prescribe mechanically the context or practice of every teacher/student or supervisor/supervisee interaction. Thus, STEP, which some would call a "cookie-cutter" approach, did not reduce the need for creativity, discretion and judgment in dealing with individual students and problems that arise regularly in human service programs. In fact, it promoted that creativity, judgment and discretion by providing a developed structure, content and approach; by giving training; and by interacting extensively with the sites throughout development and implementation.

In short, the STEP experience offers an example of innovation development that is more complex and costly than the usual central/local paradigm. The STEP approach was productive in terms of consistent application and results, and it was less divisive among central and local actors. In addition, the degree of specificity and process of imlementation allowed STEP to become institutionalized at the sites. STEP continues to operate at all five demonstration sites, three years after demonstration operations formally ended.

The STEP experience in managing program uniformity and variation, and the consistent results it achieved, also suggest the importance of developing uniformity--not only regarding program structure and content, but also staff structure and staff characteristics in those areas of the test intervention where uniformity of implementation is deemed vital. Demonstrations have, to date, primarily focused on program structure in their requirements; the frequent variation in implementation and results caused by varying content, staffing levels and idiosyncratic execution of tasks has led to concern over the usefulness and validity of aggregating data.

Allowing variation on so many operational variables assumes that central/local functional responsibilities split cleanly between conceptual design and implementation, rather than along the more complex lines that emerged in the STEP demonstration. It assumes,

and inevitably produces, highly variable results from the same design. STEP shows that a confusing variety of implementation practice, and of results, is not inevitable.

The uniformity/variation issue as experienced in STEP not only diverged from common wisdom and practice, but also prompted a major change in the research demonstration itself. STEP was originally structured to have two waves of enrollment, or cohorts, in order to achieve the number of participants necessary for the analysis. Since STEP is a two-summer program, the original design meant that during the first cohort's second summer (1986), the sites would be enrolling a second cohort for its first summer; the second cohort would thus finish STEP at the end of the 1987 summer. The long-term research would go on through 1992, when the second cohort would have had time to finish high school.

However, sites' reaction to the CORE curriculum principles and content during the 1985 summer, along with P/PV's implementation monitoring and its analysis of participant impacts for the first summer (1985) compared to the second summer (1986), persuaded P/PV not only that the first cohort did not receive a consistent remediation experience, but that the second cohort received a substantially improved remediation experience. We became concerned that aggregating the two cohorts to get program impacts would not present a fair picture of what the STEP model could accomplish.

We recommended to the funders that a third cohort of youth be added, to receive the same highly structured program as the second cohort. The second and third cohorts would then be combined to form the test sample; the first cohort would be treated as an extension of the pilot phase. This would add another year to the sites' participation in the demonstration and would extend the research to 1993. The major cost increase would be borne by the local sites, whose incremental operating costs as noted earlier, were substantially beyond the incentive money provided by P/PV.

The funders and all five demonstration sites agreed to this change, indicating the degree to which they were attuned to the needs of the experimental approach to social change. It also shows how much easier it is to perfect or adjust the mechanics of the testing process than it is to construct a substantive intervention that can deal with the many problems and disadvantages of poor youth.

Key Operating Issues

The STEP demonstration witnessed the usual implementation challenges that face any social program or, indeed, a start-up venture in almost any field. Hiring and training staff, devising and installing all the necessary operating procedures, securing the necessary funds, ensuring that the costs of new substantive elements are properly accounted for--these and many other implementation tasks consumed local sites' time and resources, were largely resolved, and resulted in sound local programs.

STEP was not unusual in this regard, nor did it produce major lessons that would help to ease these start-up challenges in future programs, beyond what was learned about the central/local

4. Although there are no accurate comparative figures available, the amount spent on the central development and implementation of specific innovations under new federal legislation like JTPA and JOBS is nowhere near STEP's.

relationship in developing and installing innovations. However, STEP did have to deal with several issues that have consistently arisen in other demonstrations and in national programs for at-risk populations. They deserve to be noted, for they raise important questions about the design and operation of social interventions in general.

Enrollment and Outreach

The pilot experience had indicated the difficulty of enrolling substantial numbers of youth--in that case, 500 14- and 15-year-olds. Of the 500, only 250 actually received STEP treatment; the other 250 were controls, who went into full-time summer jobs. The demonstration sites had the same difficulty. The demonstration requirements were that each site enroll 330 youth for each cohort, of whom 165 would be STEP participants. For all sites, it required considerable effort and resources to get 330 participants enrolled in each of the three summers of the demonstration--even when they could identify 1,000 to 1,500 youth who were academically and economically eligible. Sites had to do specially targeted mailings and intensive in-school recruitment to meet enrollment goals.

During the demonstration, the second-summer return rate for STEP first-summer completers was 75 percent. To achieve that rate required strong and intensive outreach efforts, well beyond the scope of most youth programs. P/PV site coordinators, for example, routinely called participants at home (sometimes weekly) and made home visits to persuade them to reenroll.

Simply put, youth did not flock to the STEP program. Although there is no reason to expect STEP to attract all eligible youth, the low level of jobs and remediation opportunities for 14- to 15-year-olds and the fact that STEP paid for both work and remediation made it reasonable to assume that the required 330--only a modest share of eligibles--would enroll quickly. In fact, only about half in each site enrolled in response to announcements about the availability of STEP; the remaining half required extensive, personalized outreach.

Enrollment has been a problem for many special interventions for at-risk populations. The issue receives little discussion because for both special demonstrations and most national social programs, the funds available are too small to offer services to more than a small share of the eligibles. For example, the nation's major employment and training program, the Job Training Partnership Act, has funds for between 5 and 10 percent of its eligible population. The JOBS program for AFDC recipients has statewide participation requirements that are currently 7 percent of the eligible population and will rise to 11 percent in 1992. Most states have had trouble meeting the 7 percent requirement.

That personalized outreach should be necessary to secure even modest enrollment in special programs is consistent with an increasing body of research findings about the low level of positive adult and institutional influence in many poor youth's lives, and the growing distance--both mental and physical--for many of these youth from participation in socially approved activities. But active outreach is rarely a well-developed part of social interventions. It is also one of the first expenses to be cut when program funds are reduced; from an administrative viewpoint, this type of cut is unquestionably sensible, for reduced funding usually means fewer people can be served anyway.

While calls for more program money spent on outreach seem appropriate when considered in relation to the large number of unserved eligibles, evidence

indicates that only modest levels of eligibles will voluntarily avail themselves of social interventions no matter how intensive the recruitment effort. The consistency of the weak enrollment phenomenon raises concerns about how special interventions are perceived by the eligible populations, and about the level of naturally available external or internal motivation for at-risk populations to seek out assistance in areas of personal need.

Neither STEP nor other programs offer insight into the relative strength of these factors. STEP does suggest, however, that active and personalized outreach is necessary if more than modest shares of eligibles are to participate in special interventions. The boundaries of active outreach are not clarified by the STEP experience because of the limited participation goals set for each site.

The outreach function was critical for STEP, and no doubt is and will be so for other intervention programs. The outreach capacity required of the sites, for new enrollees alone, cost about $11,000 per site per year--or around $40 per enrollee--and was supported through site grants that P/PV made periodically throughout the demonstration. Given the fact that STEP generated such consistently good short-term learning impacts, it seems important to reassess our views toward the costs of outreach and other personalized approaches designed to attract youth to self-improvement activities.

Continuity and Coordination

Both the sites and P/PV often refer to STEP as a 15-month program--two summers and an intervening school-year support component (SYS). That characterization is accurate in terms of the total time period STEP covers; however, it turned out not to be accurate in terms of continuity and coordination of program content and approach.

The STEP school-year support component was in all demonstration sites a modest group of activities that had no substantive connection to the summer's remediation, life skills or work activities. It consisted of telephone contact and primarily recreational activities, occurring monthly or less often. In the two demonstration sites (Portland and Fresno) where school-year support activities were somewhat more intensive, the additional activities were primarily advocacy and referral.

Participation rates for school-year support were low for almost all activities in all sites except Fresno. Recreational evening activities, such as pizza parties and roller skating, were the best attended, and even they rarely attracted more than one-third of the STEP treatments. STEP's approach to school-year support did not provide for continuity between the summers and the school year. Neither STEP nor the schools communicated extensively about a youth's performance and needs, and what each could do to build on what had been previously achieved by the other. Continuity and coordination between schools and employment and training agencies, between summer and school-year programming, and between school-year STEP activities and the larger culture of the school, were almost nonexistent. It seems likely that most youth saw little continuity to STEP.

In fact, what may have been most notable from the youth's perspective was the **contrast**--the disjuncture--between the STEP summer and the regular school year. Both the *Practical Academics* and *Life Skills and Opportunities* components were highly interactive, topical and individually paced. They were integrated with income-producing work on a daily basis. Thus, the STEP summer experience for many youth was substantially different in form and content from their regular

school-year experience. Since STEP youth had already had poor or failing school experiences, this discontinuity and lack of coordination may have worked against STEP's school retention goal.

Thus, STEP was a 15-month intervention of highly varying content, approach and youth involvement: two intense, high-participation six- to eight-week periods, separated by nine months of low involvement and infrequent STEP activities, and the regular school-year educational experience. This in itself would not necessarily undercut the program's effectiveness or impact; there is no evidence in the research literature or in intervention theory that would dictate consistent intensiveness. However, the extended low-involvement period during STEP was in school--the exact setting where the STEP youth had exhibited previous performance problems. Thus, a lack of continuity and coordination in its program structure limited STEP's ability to build off its summer achievements.

These shortcomings in continuity and coordination are not peculiar to STEP. Other demonstrations have had similar deficiencies. Even when special techniques like case management have been included, the operational results have been modest, as have been the reported outcomes and impacts.

Thus, STEP confirms the inability of the usual demonstration to cross institutional boundaries and achieve substantive continuity for the benefit of youth. STEP, like the two other examples mentioned above, required high-level statements of intent to coordinate from all the institutions--both schools and employment and training agencies--in each participating locality. Yet this coordination rarely went beyond minimum administrative and information requirements. Even in those sites where the schools actually operated STEP, the core school-year experience was unaffected, and neither STEP nor regular school practice built off the other's experience to improve youthful learning and behavior.

STEP did not prescribe or even suggest substantive continuity or coordination between the summer and school year. Thus, the lessons to be drawn from its lack are meager. It is notable that such continuity and coordination did not develop or evolve naturally at the local sites. Major youth-serving institutions were unable to move beyond their own core programs to seek benefits for the youth they all served.

Early Impacts

Implementation issues did not deter the successful operation of STEP's summer program of work, remediation and life skills. The first summer test score impacts were impressive for the pilot and first cohort groups, and even more impressive for the second cohort, who experienced STEP's more structured remediation component, *Practical Academics*. The third cohort's first summer results were also impressive, though they did not sustain the expectations raised by the second cohort.

Thus, STEP's early research results were consistent and positive. To fully understand these program impacts, it is useful to have a profile of the youth who participated in STEP. These were youth who were clearly at risk of dropping out of school and becoming teen parents. Figure III.2 presents the participant profile of the 2,519 youth in the STEP analysis sample.[5] All were 14 or

5. Throughout this report, we present results for Cohorts II and III, without including the first cohort, because this sample best represents the effects of STEP as a "steady-state" program.

Figure III.2 - STEP Participant Profile

I. DEMOGRAPHICS *(N = 2519)*

Age	
≤ 14 Years Old	57%
≥ 15 Years Old	43
Gender	
Male	48
Female	52
Race/Ethnicity	
Asian	19
Black	49
Hispanic	18
White/Other	14

II. ECONOMIC CHARACTERISTICS

JTPA Eligible	100%
Percentage from Female-Headed Households	51

III. EDUCATIONAL ABILITY

Mean MAT Reading Score	
Grade Equivalent	6.1
Mean MAT Math Score	
Grade Equivalent	7.0

IV. SCHOOL VARIABLES

Grade in School	
8th Grade and Below	60%
9th Grade	36
10th Grade and Above	4
Percentage Repeating a Grade	32
Percentage Experiencing Difficulty with English	15

V. LSO VARIABLES

Sexually Experienced	43%
Recent Unprotected Sex	35

NOTE: Represents the analysis sample for treatment youth in Cohorts II and III.

15 years old at program start and met JTPA low-income standards. Approximately 86 percent were ethnic minorities, and about half lived in female-headed households.

They exhibited severe educational deficits. Nearly one-third had been held back at least one grade, and all were performing below grade level in math and/or reading. (However, to ensure that youth could benefit from STEP, they had to have at least a fourth-grade reading level to be eligible for the program.) Their school records reveal sporadic attendance.

The group was approximately half male and half female. Over 40 percent of these 14- and 15-year-olds were already sexually experienced at program start, and over 35 percent of the sexually active youth had recent unprotected sex. They were clearly at risk of adolescent pregnancy.

Over the summer, STEP significantly boosted these youth's academic competence and knowledge of responsible sexual behavior. (See Figure III.3.) We found consistent aggregate positive impacts for the first summer experience of all cohorts. The results hold true for all subgroups (race and sex) and sites.

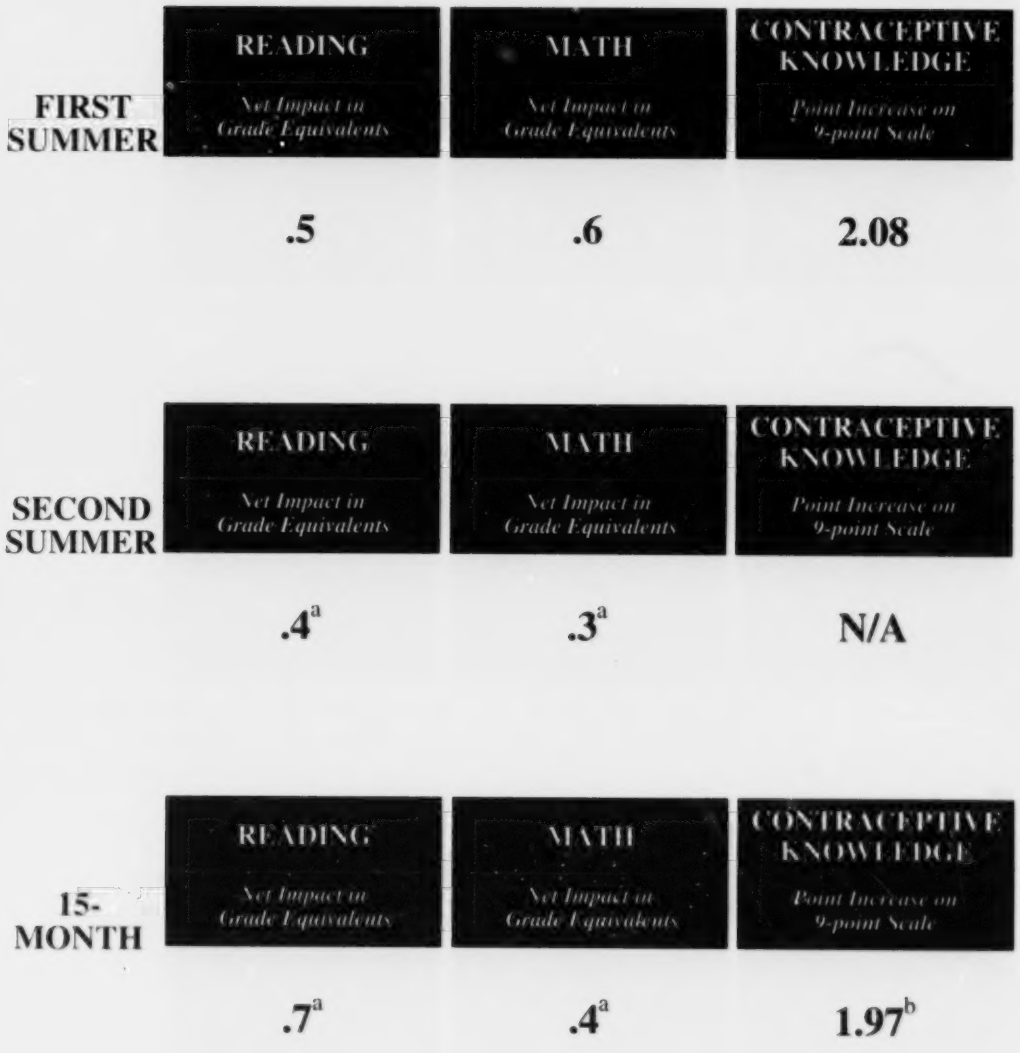
STEP's most convincing effects were in stemming summer learning loss. Participants outperformed control youth in reading and math, using the Metropolitan Achievement Test (MAT). While participants generally showed modest positive gains in academic testscores, the net impact findings are largely driven by the declines experienced by control group members. The size of the controls' learning losses varied somewhat by subject, site and summer, but the pattern is clear and consistent. Had they not participated in the program, STEP youth presumably would have experienced similar losses.

After strong first-summer outcomes, the performance of STEP youth relative to controls during the subsequent school year was disappointingly flat. We found no overall impact on reading and math test scores, credits earned or dropout behavior. When we compare the average learning gains of treatment and control youth (in Seattle and San Diego) over that school year, we find that controls gain at a faster rate than treatment youth. An analysis of school-year support (SYS) participation showed that youth who received some SYS services earned more credits and were more likely to be promoted. However, we are unable to definitely conclude that it is SYS that helped students get through school; it is equally possible that those youth who participated in SYS were in school more often and would have done better even in the absence of SYS services.

In other words, the gains of the first summer, during STEP's intense period of intervention, did not translate into changes over the school year, when STEP was barely present. As noted earlier, youth returned to the same school environment in which they were failing before STEP. There was no substantive continuity or coordination between the STEP summer and school-year activities. Whatever gains may have been made over the summer in learning, skills or motivation evaporated in the face of a school experience where nothing had changed and no attempt was made to build off the success of the summer experience.

The second-summer effects are less strong than those of the first summer. In both academic tests and measures of knowledge of contraception, results continue to show a positive net impact of participating in STEP. However, treatment youth gained less than in the first summer, while control youth in the two

Figure III.3 - STEP In-Program Impacts



NOTE: The sample sizes used in calculating the above impacts are as follows. First summer reading and mathematics, 2511; contraceptive knowledge, 2082. Second summer reading, 617; mathematics, 592. 15-month reading, 676; mathematics, 677; contraceptive knowledge, 2501.

[a] These impacts are based on the data in San Diego and Seattle only.

[b] These impacts were observed at 18 months.

sites where they were tested continued to experience losses. Why would the second-summer impacts have fallen off? Perhaps the disappointment of the intervening school-year performance affected the youth's view of STEP's usefulness; we know it took much follow-up and persuasion to convince a high number of participants to return for a second summer, and the smaller gains may reflect a lower level of motivation and interest, or boredom with an experience that was too similar to that of the previous summer.

When we put this evidence together, we find that participation in the full 15-month STEP program improved the reading and math scores of participants relative to controls, primarily because the summer impacts in combination were so substantial. During the school year, controls gained at a somewhat faster rate than STEP youth. The differential learning rates between treatment and control youth during the school year, as discussed earlier, could be attributed to disappointment on the part of treatment youth with regular school. It is also possible that schools are prepared to handle summer learning loss by spending the first month or six weeks of each year compensating, by review or slow movement through new academic material. STEP youth who had not experienced that loss would not gain during this period, thus completing the year essentially level with their peers.

As for STEP's impacts on teenage pregnancy and parenting rates, effects on actual behavior are not likely to be measurable within the relatively short term of the intervention (15 months). However, changes in contraceptive knowledge, sexual activity and knowledge of the consequences of teen parenting could occur within that period. And in fact, STEP youth did demonstrate significant improvement in fertility-related attitudes and knowledge compared with their control group peers. Behavior, however, is more difficult to modify. While we found encouraging glimmers among certain subgroups or in one or the other cohort--e.g., a 53 percent increase in contraceptive use among sexually active treatments in Cohort II--we were unable to detect a consistent pattern of in-program change in social or sexual behavior.

Overall, the in-program impacts leave the impression of a program that is achieving in the short run exactly what it intended--the prevention of summer learning loss in reading and math, and gains in knowledge and understanding about real-life issues involving sex, health and career. In addition, the research surveys indicated that youth rated their work experience very highly--twothirds had not previously worked, and felt they learned a great deal about responsibility, workplace expectations and practices. All early indications were that the program was effective within its limited, intensive summer periods, though no indications had yet emerged of effects on school-year performance.

IV. Replicating STEP

The STEP demonstration's consistent and positive summer results brought the STEP model to the attention of policymakers and administrators. National interest in dropout prevention and educational improvement was intense, and the debate was focused on the critical importance of basic skills for young people if they and the American economy were to prosper. In addition, the 1986 JTPA amendments mandated the inclusion of remediation in the summer jobs program for all youth who needed such assistance. Early local response to this law was weak; local operators complained that they did not have the resources and expertise to implement the required remediation. As a result, employment and training officials at the national, state and local levels felt that STEP's strong early showing was sufficient to justify its use in their programming.

Another issue pushed STEP's replication: the lack of success in replicating previous demonstration models, and lack of knowledge about the replication process itself. Lack of replication experience was producing a tenuous connection between research demonstration work and actual practice in the field. Some policy leaders believed that effective interventions were a function of unusually talented local leadership--that the problem was not a lack of effective program interventions, but a shortage of local leadership, which imposed a severe limit on the extent to which replication could occur. Only large-scale replication of a complex, innovative model like STEP could provide experience and information. Replicating STEP, a well-developed program with clear evidence of short-term success, would allow an exploration of the potential and mechanisms of large-scale installation. The existence of short-term demonstration impacts provided a standard by which to assess the effectiveness of the replication.

The Replication Strategy

Planning for a STEP replication initiative started in 1987, when the second cohort's first-summer results, which were impressive, became known. With support from the U.S. Department of Labor and the General Motors Corporation, P/PV initiated the first phase of replication in Summer 1988, in four states and 11 sites. This first effort was cautious and small, acknowledging the lack of experience and knowledge about replication.

Later in 1988, encouraged by the success of this first wave of replication and anxious to assist a larger number of sites in the development of summer programs that complied with the mandate of the JTPA amendments, the Department of Labor funded P/PV to carry out a larger replication effort.[6] In 1989, STEP replication also secured a second corporate sponsor, the Exxon Corporation. Thus, in Summer 1989, replication increased to nine states and 37 sites. In 1990 and 1991, STEP replication was further expanded, bringing STEP to 15 states, more than 100 sites, and 8,000 youth. (See Figure IV.1.) To date, more than 20,000 youth have been served by the replication sites.

The first replication effort in 1988 provided the prototype strategy for the succeeding phases. Although the increasing size and scope of each phase and the increasing level of experience of many sites in operating STEP caused some changes in the replication strategy, the 1988 plan has held basically constant. That plan had four central components:

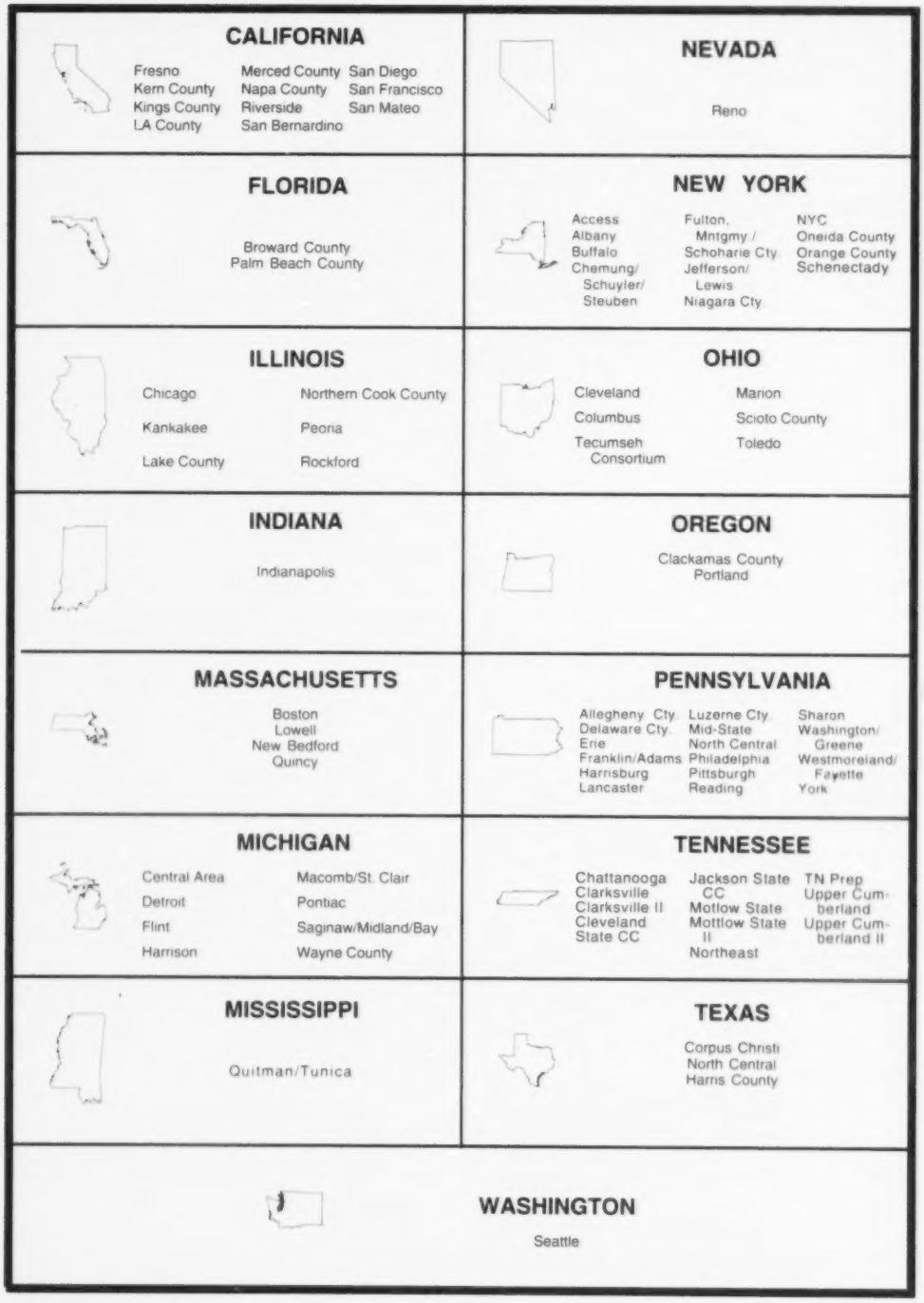
1. **A State-Focused Marketing Strategy**. Our intention from the beginning was to replicate STEP in a manner that might later promote further replication without intensive assistance from P/PV. It seemed unlikely that this could be achieved by training staff at the local level, where both the numbers of staff and discretionary resources were limited. At best, those people might be able to help expand staff within their local jurisdictions, but they probably would not be able to provide training to sites in other localities.

   The state level seemed to be the right market choice for long-term consideration. Under JTPA, states control discretionary monies for special programming, training and technical assistance and take responsibility for generating statewide plans regarding coordination of JTPA with other agencies, such as the state education department. Also, since state education departments have considerable influence over local education (over the past decade, states have represented the fastest-growing source of financial support for public schools), a state-level marketing strategy for STEP replication seemed to offer potential for greater employment and training/public school coordination.

   Thus, we fashioned a marketing and selection process aimed at the states. We enlisted the assistance of the National Governors Association in identifying and contacting states that might be interested. Our proposal required a participating state to commit to work with P/PV and, initially, with at least two or

6. In 1988, the Rockefeller Foundation underwrote our initial exploration of the issue of replication in social programs. This work has been carried out in tandem with the replication of STEP.

Figure IV.1 - The National STEP Network



BEST COPY AVAILABLE

three localities. The states were required to name a state-level staff person as the coordinator of the effort. In addition, we asked the states to cover half of P/PV's training and technical assistance costs (the other half was covered by P/PV's national DOL grant), and to provide further unspecified assistance to the localities. The local sites were asked to cover the incremental costs of operating STEP over the usual summer jobs program. The level of support varied, with some sites providing the P/PV-specified minimum and others going well beyond requirements. Pennsylvania, for example, sponsored almost 20 STEP sites statewide.

Having a state-level staff person committed to STEP, we hoped, would encourage further replication within the state. Replication could become part of a larger state strategy, and technical expertise would be in place at the state level. To promote an active state role that utilized P/PV's experience, we devised a state-managed process in which P/PV helped states solicit and select localities to operate STEP. A formal agreement between the state, the locality and P/PV was established, requiring the Service Delivery Area (SDA) to implement the STEP model as designed and operated in the demonstration. Our sense was that we would probably not be able to retain programmatic control over STEP if its replication grew large, but that in the initial stages, we should try to ensure fidelity to the model that was producing impressive short-term impacts.

2. **A Reduced-Cost, Multiple-Source Financing Strategy**. Given the JTPA system's lack of experience with, and dedicated budget lines for, replication activities, we knew that central oversight, training and technical assistance costs would have to be reduced from their demonstration level of about $33,000 per year per site for three years. The final plan was for two years of assistance to each site, at about $15,000 per year. The cost savings were achieved by deleting the use of on-site central staff, reducing monitoring visits, and relying more on training conferences and specialized assistance. The concern at P/PV was about not reducing the level of assistance to the point where we could not assume fidelity and consistency of operation and results. The countervailing consideration was that three years of experience provided a base on which to shape a more efficient and less costly installation strategy.

Even this reduced amount was judged by those familiar with JTPA to be too high to market to the JTPA system alone; they did not believe the system would be willing or able to bear the initial costs of learning to implement STEP, in addition to its incremental costs. Given that STEP had not been successful in garnering financial support from the U.S. Department of Education or significant support from the local school districts, we also did not expect to get state education financial support in the initial replication stage.

Thus, the financing strategy for covering installation costs was based on multiple sources of support: philanthropic and corporate funds for marketing and early development costs; an equal combination of federal and state JTPA funds to cover training and technical assistance; and local JTPA funds, supplemented in some cases by state funds, to operate the

program. The incremental operational cost of two summers of STEP over the summer jobs program was about $1,200 per participant, so even with the infusion of corporate, philanthropic, state and federal funds for development and installation, the local employment and training agency was being asked to divert a significant amount of available operating funds from other programs and participants to its STEP participants.

3. **A Streamlined Training and Technical Assistance Strategy**. The demonstration approach to installing and maintaining STEP had involved not only P/PV staff and specialized consultants, but also a part-time, on-site staff person, reporting to P/PV, who worked with the sites on program, research, management and information issues that arose during the course of the demonstration. P/PV staff and remediation and LSO consultants also visited sites frequently during the spring planning and summer operational phase. Thus, assistance had been hands-on and intensive.

   Reducing installation costs for the replication meant that there had to be a less-intensive approach. The plan included national and statewide training conferences; extensive, well-designed and frequently updated manuals and curricula to codify all the aspects of operating the program that had arisen during the demonstration; and the use of a group of consultants who had been involved in the STEP demonstration and were located in various regions of the country. Each consultant was assigned to work with a group of states and local sites, and was available for several on-site trips to local programs and limitless consultation by phone.

4. **A Quality Control/Performance Feedback System**. In the pilot and all three demonstration cohorts, the STEP model had proved its capacity to produce consistent and positive summer impacts. The test score gains of STEP participants in both educational and life skill areas would provide benchmarks for the replication programs, so that they could know if their results were equivalent to those of the demonstration.

   Thus, we provided testing procedures and training to each site, and had them mail all tests directly to P/PV. We then processed the results and provided them to the states, local operators and national funders. This testing and feedback operation added to both local and central replication costs, but it was judged from the outset to be essential if sites were going to meet the rigorous standards of replicating not just STEP's ideas or principles, but its results as well.

It was, of course, critical to have a sound strategy for replicating STEP. A poorly conceived strategy regarding marketing, costs or financing might have led to limited local interest in operating STEP; if the training, technical assistance and materials approaches were inadequate or poorly executed, STEP replication sites might have generated less impressive short-term results.

An equally critical issue was deciding which parts of the STEP model should be replicated. The most impressive results were from the first-summer component. The second-summer results for the first cohort were less impressive, and both observation and test scores indicated that the school-year support component was achieving little in relation to school performance. There was discussion about assisting sites who

only wanted to operate a one-summer program or who wanted to construct an entirely different second-summer experience.

The final decision was to replicate STEP only in sites that would operate the full STEP model as designed. Since the long-term results were not in, it seemed unwise to let the existing model be diluted or significantly changed until it was shown whether it produced long-term impacts. The national replication funders agreed with this decision; they, like P/PV, were also interested in the broader question of whether a complex model like STEP could be faithfully replicated in many locations and consistently produce the same program effects that the demonstration had shown to date.

The content of the work experience and school year support components are, as in the demonstration, left to local discretion. However, the operating handbook emphasizes more strongly the importance of the school-year component and offers examples of ways the replication sites could improve on the demonstration experience. Since replication sites receive no financial assistance from P/PV and have no on-site P/PV staff presence, central control over what sites actually do with their school-year component is minimal.

Results

The replication initiative has not included random assignment or long-term follow-up of participants. It has not been carried out as a research or demonstration effort, but rather as a businesslike effort to install the model with consistency in many locations, at the lowest feasible cost. As noted earlier, it has included formal testing on both educational and life skill measures, and collection of data regarding participant characteristics and participation.

The replication effort cannot be judged solely, however, on its ability to generate participation and test scores similar to the demonstration's. Since it is aimed at installing STEP more widely, it also raises issues of penetration, coverage and viability across a wide variety of locations.

Penetration and Coverage

In 1991, 15 states participated in STEP replication. All 50 states have received materials and been approached during STEP's four-year replication effort, and no state willing to accept STEP's programmatic and financial requirements has been turned down for replication assistance. Thus, this one-third state penetration rate probably represents very near the upper limit of states willing to take on STEP replication in its current form and under its current financial packaging.

No state that has begun STEP replication has decreased or abandoned the initiative. Eight of the 15 have expanded the number of replication sites; one, Pennsylvania, has made STEP a major statewide initiative; five more have also carried out substantial expansions. Thus, STEP appears to have rooted in those states and localities where it was installed, and has naturally expanded through state adoption at a moderate level. Most of the STEP states indicate that financial limitations are the primary obstacle to further expansion.

In 1991, STEP was operating in over 100 localities in 82 of JTPA's SDAs. Given that the JTPA system, for funding purposes, divides the country into 620 SDAs, the local penetration rate is 14 percent. Unlike the state penetration rate, this number does not represent a good estimate of the desire of local communities to operate STEP. As noted

earlier, most states ran competitions to select local STEP operators; although data was not kept on the number of localities that applied, P/PV staff estimate that at least 75 SDAs were rejected in the competitive process.

In addition, although the marketing effort was aimed at states and not localities, 35 SDAs have communicated directly with P/PV about operating STEP during the four years of the replication initiative. In most cases, they were in states that had decided not to respond to the initiative. In all but three cases, P/PV could not work out a satisfactory arrangement to work directly with the locality, usually for reasons of financing or central training capacity.

Based on the replication experience, the outer limit of local interest in operating STEP cannot be solidly established. It is clearly above 14 percent; how far above we have no way of knowing, though simply including those SDAs that have competed in state selection processes and contacted P/PV for help would raise the local penetration rate to over 30 percent. Only a local marketing strategy, combined with direct subsidy of local installation costs, would provide a more accurate estimate of potential local penetration. Without subsidization of installation costs, the local penetration rate would be substantially below the current 14 percent.

In 1991, the 100 STEP replication sites served about 8,000 youth, a substantial increase over the 750 treatment youth who were served in each demonstration cohort. In total, the four-year STEP replication initiative has enrolled over 20,000 youth. But while this is a substantial number of youth served, it is unlikely that it represents a substantial percentage of the total youth eligible. National statistics are not maintained that would help provide an accurate number of such youth, but a best estimate, using available statistics, is that about 700,000 youth meet STEP's eligibility criteria each year. Thus, STEP's scale of replication has been notable for a voluntary, specialized intervention initiated as a research demonstration, but it has probably provided services to only a little over 1 percent of eligible youth.

As discussed earlier, there is strong evidence that even if STEP were available to all eligible youth, far fewer than the full number would participate. We have no way of projecting the percentage of those who would participate if they could, but the considerable effort spent by demonstration sites to attract between 20 and 30 percent of eligibles probably allows a reasonable guess at upper participation limits. If so, it would mean that STEP replication has reached 3 to 5 percent of those youth who would voluntarily participate. This rate is very similar to the estimated coverage rate of nationally legislated programs such as the 1982 Job Training Partnership Act (JTPA) and the JOBS program of the 1988 Family Security Act.

In sum, STEP replication's penetration and coverage have been significant for a program begun as a research demonstration, though far below both local interest and thus student willingness to participate. The decision to insist on the full STEP model, and the cost and capacity implications of that decision for local and state decision-makers, no doubt dampened both state and local interest somewhat. In addition, the state-focused marketing strategy, though necessary for financial reasons, denied participation to many localities that were willing to finance STEP's incremental operating costs.

Another major limit on STEP's capacity to provide services in all interested communities to all willing eligibles is the location and level of funding for training and technical assistance--in short, for

replication. In the JTPA system, funds for program installation are lodged primarily at the state and federal levels; localities have very limited resources for installing a program like STEP. STEP tapped into those state and federal funds to cover most of its local installation costs; even then, however, it was necessary to raise philanthropic and corporate funds to underwrite marketing and materials development. The total cost to carry out replication installation from 1988 to the present has been close to $5 million: 29 percent came from the U.S. Department of Labor; 65 percent from states; and 6 percent from philanthropies and corporations. The complexity of developing and maintaining this financing package imposes another serious restraint on STEP's potential--and probably that of any other replication initiative--for maximum penetration and coverage.

Program Effectiveness

Unlike the STEP demonstration, STEP replication has no formal research component. There is to be no long-term follow-up, no control youth, and no P/PV or local staff dedicated to observing and analyzing local operations. The information collected in the STEP replication initiative is aimed at ensuring that only eligibles enroll and seeing whether learning gains are achieved.

This limited information-collection strategy was established primarily out of the desire to use the national funds that were raised to expand STEP, not for further research. We were confident that states and localities would not have been willing to absorb higher costs for research purposes, given the costs they were already being asked to bear. As a consequence, the replication initiative has not produced a detailed body of information about program operations. It has, however, produced basic information about the youth who enrolled and their summer test gains.

The Youth. As was true during the STEP demonstration, replication sites have thus far attracted a high proportion of minorities: 47 percent were black, 14 percent were Hispanic, 28 percent were white, and 11 percent were "other." Thirty-nine percent of the youth said their family received AFDC or other welfare payments. All were poor.

About 51 percent of the enrollees reported having failed or repeated one or more grades in school. On average, they read at the fifth- to sixth-grade level--typically three years behind their actual grade levels. Almost three-quarters were sexually experienced, and almost half had recently had unprotected sex.

Thus, STEP replication enrollees, like their demonstration counterparts, were already exhibiting serious signs of poor school performance; since most were sexually active, they were beyond the purely preventive aspects of the life skills classes. STEP's modest intervention approach once again attracted a group of youth at substantial risk of not completing high school and of not being able to compete successfully for decent-paying jobs. Figure IV.2 presents a profile of program participants.

Completion rate. Seventy-five percent of the youth in the five demonstration sites who enrolled the first summer also enrolled in the second summer. In contrast, slightly more than half of early replication enrollees who enrolled for the first summer enrolled for the entire program. Our initial reaction was one of surprise at the serious decline in program completers. However, analysis of demonstration implementation practices indicated that the most likely explanation was a limited personalized outreach effort by the replication sites to encourage youth to reenroll for the second summer. The demonstration sites, at P/PV's urging and with its financial

Figure IV.2 - STEP Replication Participant Profile

I. DEMOGRAPHICS *(N = 5697)*

Age	
≤ 14 Years Old	58%
≥ 15 Years Old	42
Gender	
Male	54
Female	46
Race/Ethnicity	
Black	47
Hispanic	14
White	28
Other	11

II. ECONOMIC CHARACTERISTICS *(N = 5697)*

JTPA Eligible	100%
Percentage Receiving Welfare	39

III. EDUCATIONAL ABILITY *(N = 5376)*

Mean Reading Score	
Grade Equivalent	5.5
Mean Math Score	
Grade Equivalent	6.0

IV. SCHOOL VARIABLES *(N = 5376)*

Grade in School	
7th Grade or Below	18%
8th Grade	56
9th Grade	22
10th Grade and Above	4
Percentage Repeating a Grade	51

V. LSO VARIABLES *(N = 2564)*

Sexually Experienced	72%
Recent Unprotected Sex	45

support, had devoted extensive efforts to reenrolling youth. While STEP was being tested, it was important to see that as many treatments as possible received the full STEP intervention. Reenrollment had been a priority operational issue during the demonstration, and P/PV site coordinators played a significant role in carrying this out.

But, as discussed earlier, public programs with funds sufficient to cover only a small percentage of eligibles are typically not willing to devote resources to outreach and to other functions that require extensive and costly personal contact. In addition, the JTPA system has not funded or created many services that cover a long time span, as STEP does. Thus, it does not have extensive experience or capabilities for outreach to maintain participation.

The replication initiative's modest and marginal school-year support component, combined with minimal outreach, produced a "natural" completion rate of 50 percent--that is, one in two STEP students were motivated enough by the first summer's program to want a second, similar experience. During subsequent replication years, sites were informed of the need for extensive outreach to keep students involved. As a result, the reenrollment rate in Phase III replication sites increased, but did not reach the same level as the demonstration sites.

Knowledge gains. It is not possible to directly compare replication and demonstration program results. The demonstration research followed an experimental design with controls and treatments, and imposed on the sites use of the Metropolitan Achievement Test (MAT) to measure in-program academic gains. The replication initiative had a limited evaluation component that did not include a control group. In addition, P/PV did not carry out or finance the replication sites' tests or their administration--only their scoring and reporting. Thus, P/PV felt it could not impose, but could only suggest, the use of the MAT. Sites were given the option of using the MAT or the Test of Adult Basic Education (TABE), which is more familiar to employment and training operators. Most replication sites have chosen to use the TABE.

However, comparisons between scores can be constructed. There are accepted procedures by which test scores on different standardized tests can be converted to grade-level gains or losses. Using those procedures, we are able to make rough comparisons. As shown in Figure IV.3, the replication sites had consistently impressive reading, math and contraceptive knowledge gains. In fact, converted to grade-levels, the gains by replication youth exceed those of treatment youth in the demonstration. This could be a result of imperfect conversion formulas; the use of different tests; improved training techniques and materials; programming; or the smaller second-summer return rate--if we assume students who are performing poorly are less likely to return for the second summer.

Whatever the explanation for replication's apparent improvement in academic gains over the demonstration experience, more important is the consistency of measured academic skills improvement in both the demonstration and replication efforts across a wide number of geographically dispersed and widely diverse locations. STEP's demonstration results were consistently replicated on a significant scale. In short, replication has worked.

Figure IV.3 - STEP Replication In-Program Effects

NOTE: The sample sizes used in calculating the above impacts are as follows: first summer reading and mathematics, 4067; 15-month reading and mathematics, 1987; first summer contraceptive knowledge, 2065; 15-month contraceptive knowledge, 2065.

Implementation Lessons

As noted earlier, the replication initiative has not had a research component. Thus, we could not systematically collect or analyze implementation events or patterns. In addition, on-site observation was infrequent. Multisite training sessions and requests for assistance produced most of the knowledge of site operations. Nonetheless, these sources of information produced enough knowledge about local operations to suggest that the STEP demonstration's major operational features and themes were repeated in the replication and to generate information that advances our knowledge about the practice of replication.

Central/Local Relationships

The "cookie-cutter" versus "local creativity" debate was, once again, not a major issue. With few exceptions, the replication sites welcomed a well-packaged program structure and components, detailed training sessions and operational manuals.

One problem did arise as new replication sites were added each year: local staff came to each year's national training conference at different levels of experience. To contain central costs, establish an active network, and promote interchange and learning, different cohorts of sites do not have separate training conferences. However, the annual training conference now includes several tracks that provide opportunities for new and experienced STEP operators to meet separately.

Nevertheless, the larger lesson holds: the local/central debate over programming content subsides considerably when there are sufficient central funds to provide training, materials and frequent updates of them; and competent, responsive assistance to help resolve local problems.

Program Consistency

The second major operational theme of the demonstration that was confirmed in the replication initiative was that it is possible to generate highly consistent implementation practices across multiple sites. The demonstration had only five sites, carefully selected and supported with funds and on-site assistance, so their operational consistency proves little. But there are 100 replication sites; the state selection processes are necessarily less restrictive than the demonstration process; no central funds are supplied to local operations; and the training and technical assistance provided is a careful balance between uniform training sessions for all sites and individualized responsive assistance, in order to reduce per-site installation costs. This situation provides a reasonable test of the potential of replication as a technique for producing consistent implementation.

The student test results indicate consistency of implementation, as do nonsystematic observation and anecdotal evidence. To provide direct evidence, P/PV assembled a team of its own staff who had worked on STEP, STEP consultants, and The Ford Foundation official who had originated the STEP idea--none of whom had any role in the replication. Nine of the current replication sites were randomly selected for observation, and an observation instrument was developed.

All nine reports indicated that STEP replication sites were being operated according to the original STEP model; seven of the reports judged the STEP replication site to be superior in operation to the demonstration sites the writers had observed.

We suspect that this consistency of operation is a function of the experienced

staff and consultants who provide training and other assistance to the replication sites. And although the per-site installation costs of replication are significantly below the per-site costs of the demonstration, they are still substantially higher than the typical costs of installing model programs. The most widely used techniques in the social policy world are one-time training sessions and the provision of operational manuals; ongoing training, on-site specialized assistance and quality control/performance feedback systems are rare.

A related theme suggested by the STEP replication experience is that it does not require exceptional leaders or a rare breed of human being to implement operationally complex human service programs. The STEP replication effort produced highly similar operations and results in 100 locations across the nation; it is unlikely that all of the staff involved were unique or exceptional. What does seem clear is that competent and committed management and service staff are more common at the operating level than many social program critics allow. The replication record of STEP should provide hope that **operating** effective social interventions is not a rare or mysterious talent, but more a matter of adequate resources, defined strategy and content, and guidance and ready assistance regarding concrete implementation issues and problems.

Continuity and Coordination

During the replication effort, P/PV has communicated to sites the importance of having a strong relationship between summer and school-year programming, and has generated materials and conducted workshops on the issue. The requirements for the STEP program have not, however, increased to include particular models of school-year programming. Thus, the content of STEP during the school year has depended largely on local discretion and inventiveness.

The resulting levels of substantive continuity and coordination have been a repeat of the demonstration experience. Because of the large number of replication sites, there are, of course, a number of exceptional efforts to establish program continuity and coordination with school-year programming. But these efforts remain exceptions; even they have been unable to bring STEP's remediation or *LSO* approaches into the regular school day for STEP youth.

Thus, the demonstration pattern has prevailed: STEP is, for the most part, a discrete, effective summer program that from the youth's perspective has little relationship to his or her schooling. We would be surprised if the school-year performance of STEP replication's students was substantially improved over that of STEP's demonstration students.

Future Directions

STEP's successful wide-scale replication provides important information about the potential of replication as a social policy tool. It shows that it is possible to install a complex, multicomponent intervention in many locations and produce consistently positive results--and not generate in localities the sense that they have been "cookie-cuttered." It also shows that uniquely outstanding (and, therefore, rare) local operational staff are not the only ones who can effectively carry out such a program. These are lessons that open new avenues for thinking about achieving more consistently effective local programs for youth.

For STEP and P/PV, continuing the replication initiative beyond its current state poses both challenges and opportunities. To increase penetration and coverage, P/PV has in the last year instituted a peer team component, so that state officials and local staff with

several years of STEP operational experience have the tools to train other local staff, and thus reduce the need for and costs of central training and assistance. Three states--Pennsylvania, Tennessee and New York--are participating. The effectiveness of this strategy won't be known for several years. Although early results appear promising, our projection is that the demands of local staff's regular jobs may impose natural limits on the use of this strategy, and that it will be a complement to, not a substitute for, an ongoing central training capacity.

Improving STEP's substantive connections to school-year programming remains a major challenge. We are approaching school systems that are undergoing or are interested in undergoing major changes to see if STEP can be more integrated into year-round programming from its inception. We are also planning to strengthen STEP's substantive requirements, to ensure that new STEP sites entering through the employment and training system see the school-year component as a major program element from the beginning.

V. Long-Term Impacts

The impressive learning impacts generated during Cohort II's first summer (1986) generated optimism among STEP's funders, administrators and operators about STEP's potential to have substantial and enduring effects on young lives. As recounted earlier, the remediation component had been substantially fleshed out for the summer of 1986, and was renamed *Practical Academics*; the Summer 1986 impacts appeared to justify that adjustment of the STEP design.

The 15-month results for the first cohort were less heartening, but did not significantly dampen enthusiasm. These results were still positive and there were strong reasons for seeing them as a floor rather than an outer limit of STEP's potential for impacts. First, the first-summer foundation for STEP was a weaker remediation component than that represented by *Practical Academics*. Second, the school-year support component was largely ignored during the planning, pilot and first cohort phases.

It was felt that steps being taken to strengthen the program would produce greater effects for participants in the second and third cohorts. *Practical Academics* was being more fully developed, and the school-year component was receiving more attention at both the central administrative and local operating levels. Guidelines and possible school-year activities were being debated and drafted, and as local employment and training agencies gained experience working with local school systems, they gained confidence in their ability to achieve program continuity and coordination across the school year.

Thus, it was decided that the second and third cohorts, whose participants would receive the strengthened summer and school-year support programming, would receive long-term follow-up. The first cohort would be treated as the

pilot cohort had been: as an early operational trial, whose major purpose was to locate and remedy "bugs" in STEP's design and operation. The addition of a cohort in mid-course, with its increased cost and extended timing implications, shows that STEP's funders, administrators, policy advisors and local operators were committed to a substantively sound research demonstration process.

Follow-up interviews were conducted in late 1990 and early 1991 for both the second and third cohorts; school records data for those youth were also collected. For Cohort II, the interview and school records data provide information on educational, fertility- and employment-related behavior approximately four and a half years (54 months) after youth's STEP enrollment, that is, three and a quarter years after their participation in STEP ended. For Cohort III youth, the data provide information on behavior approximately three and a half years (42 months) after enrollment, or two and a quarter years postprogram.

The majority of the Cohort III youth were between 17 and 18 years old when follow-up data were collected; approximately half of them would have graduated if they had completed one grade each year after entering STEP. The other half would have been in 12th grade. Cohort II youth were one year older, 18 or 19 years old, at the time of data collection; nearly all were scheduled to have graduated.

The results of the data analysis are consistent and clear: STEP had no long-term impacts on its participants. Both controls and treatments started the program with similar educational deficits (described in Chapter III), and nearly half were sexually experienced. Several years after STEP, the two groups were still similar--with regard to education, employment and parenting. The lack of long-term impacts is quite consistent across race/ethnicity, gender, site, initial grade level and cohort subgroups--much as the short-term impacts were consistent, albeit positively so.

Chapter IV (Postprogram Impacts) of the STEP *Report on Long-Term Impacts* (Grossman and Sipe, 1992) provides greater detail on the research findings. It also discusses the few anomalies and discrepancies that did arise among cohorts and other subgroups (e.g., STEP was found effective in reducing dropout behavior for those youth who entered the program following the ninth grade in Cohort II--but not in Cohort III). But these glimmers of hope, in light of the overall findings, seem just that: glimmers. The few positive subgroup and individual cohort effects do not form a pattern. Although they are interesting and will be subject to further analysis, they do not dilute the overall finding that STEP was unable to keep youth in school; to raise their grades and test scores; to delay or reduce sexual activity or pregnancies; to increase post-high school employment rates; or to reduce the need for welfare assistance. STEP was an intensive intervention that produced immediate positive effects, yet had no measurable impact on later behavior, decisions or activities.

Why STEP produced no long-term impacts is of course a pivotal question; in its answer lie the seeds for future initiatives to encourage positive change in young lives. That answer cannot be produced with the same rigor and certainty as the impact findings; we are thrown back to the same process of intelligent guessing that characterizes most program design. But before seeing what insights and lessons STEP may hold for future efforts, it is useful to see the status of STEP youth, both controls and treatments, several years after STEP.

STEP Youth In 1990

In 1990, controls and treatments looked similar across the major life variables studied. Aggregated, they are a relatively large sample of poor urban youth, and thus offer a compelling and credible portrait. In the absence of comparable national data on other urban, disadvantaged teenagers, it is impossible to say with certainty whether this portrait is characteristic of similar youth nationwide. But it is clear that many of these youth, eligible for STEP by reason of low income and poor educational attainment, display life patterns with decreasing rates of attainment over time (Figures V.1 and V.2) that raise serious concern about their ability to achieve self-sufficiency in the current and projected job market.

Education

Three and a half years after their enrollment, we find that while half of the Cohort III youth were expected to have graduated high school, only about a quarter had a high school diploma. Four and a half years post-enrollment, when all of Cohort II should have graduated, only 55 percent had diplomas. Approximately half of those who had not graduated "on time" were still enrolled in school.

About 22 percent of Cohort III had dropped out of school; and about 27 percent of Cohort II (one year older than Cohort III) had dropped out. While these dropout rates are much higher than the national average of 12.1 percent (National Center for Education Statistics, 1990), they are similar to those reported in many inner-city school systems. For example, the New York City school district reported that four years after its Class of 1988 entered high school, just over 20 percent had dropped out but another 25 percent had enrolled for a fifth year.

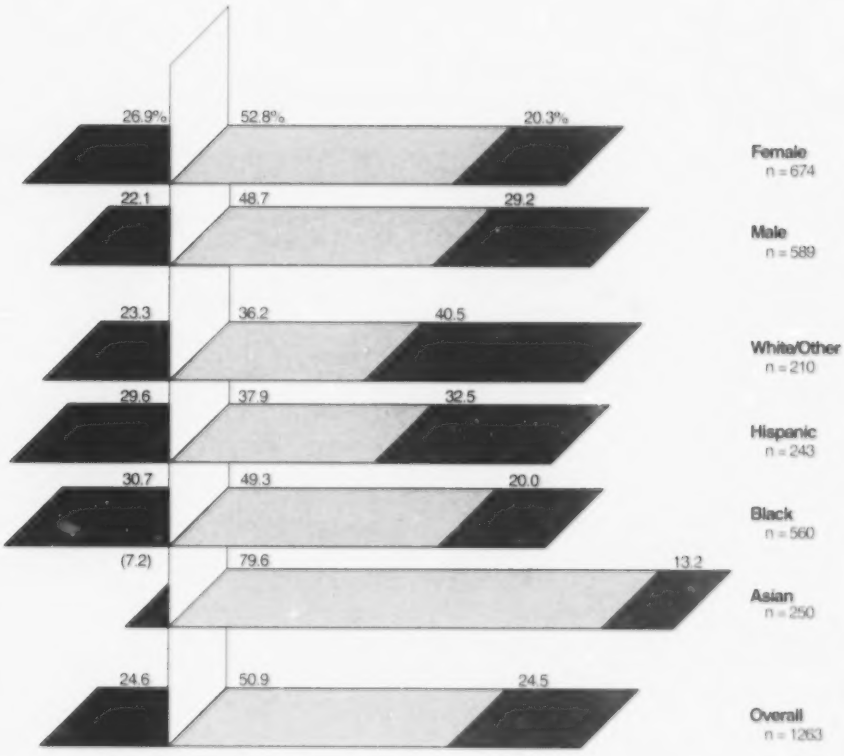
Fifteen percent of the younger Cohort III, and between 25 and 30 percent of Cohort II were enrolled in college. The proportion of STEP youth going on to college seems high, given the program's eligibility criteria. Taken as a percentage of individuals who had received a GED (5 percent) or a diploma (approximately 55 percent), this means that approximately 49 percent of Cohort II's college-eligible youth continued on to college. Using census data, Carter and Wilson (1989) report that 57.5 percent of all 18- to 24-year-olds attended college in 1988, but that among minorities, the rates were much lower: 30.3 percent among low-income black graduates nationwide and 35.3 percent among Hispanic graduates. The comparable rate among STEP Cohort II Hispanic youth was the same as the national average. However, STEP black youth were a much more college-oriented group than average: 45 percent of STEP Cohort II black graduates went to college.

In total, about 70 percent of the 17- and 18-year-old Cohort III youth were involved in some type of schooling; of the 18- and 19-year-old Cohort II group, about 50 percent were involved in schooling. Although the experiences of individual members of the two cohorts may not be identical, the 20 percent difference between the schooling rates of 17- and 18- and 18- and 19-year-olds of similar backgrounds suggests the possibility of a continuing and rapid decline in formal educational activities.

Work

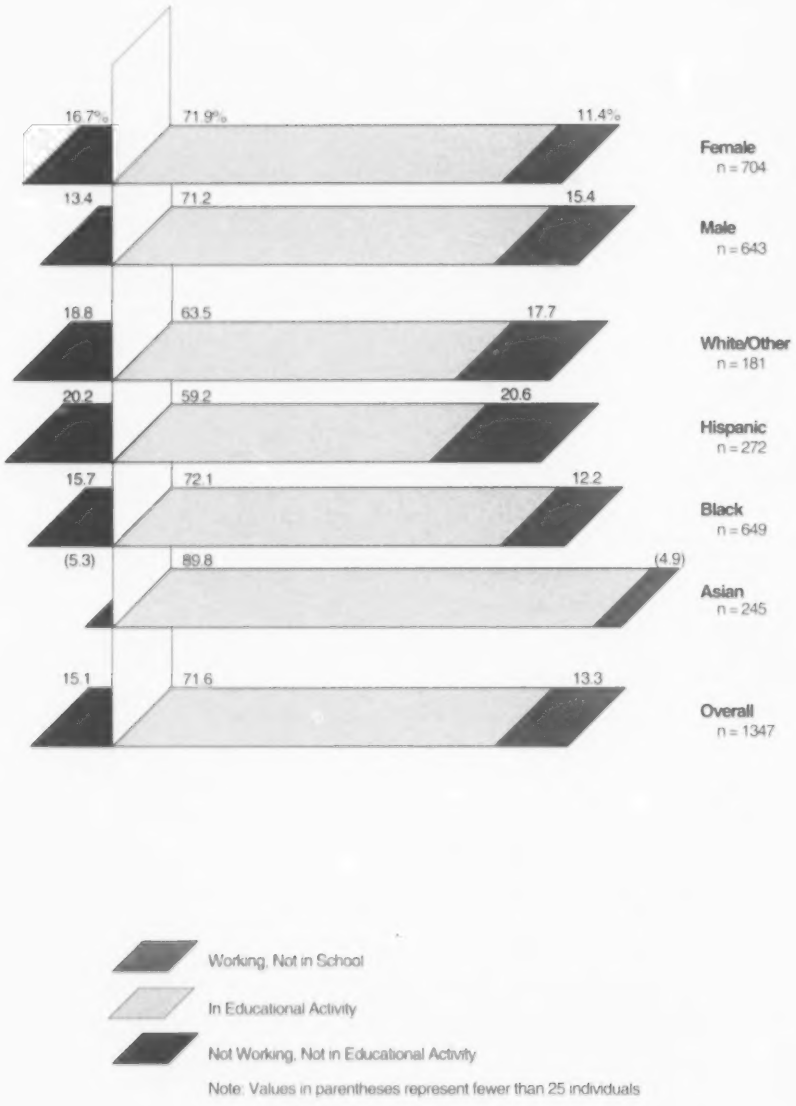
Of the youth who were not attending school, approximately half in both cohorts were working at the time of the 1990 interviews. National and STEP youth employment rates appear comparable. Nationally, 60 percent of all 17- and 18-year-old out-of-school youth were employed at some time during 1988 (Simons et al., 1991); the

Figure V.1 - Economic and Educational Activities of Cohort II STEP Youth 54 Months after Enrollment



Note: Values in parentheses represent fewer than 25 individuals

Figure V.2 - Economic and Educational Activities of Cohort III STEP Youth 42 Months after Enrollment



percentage of individuals employed some time during a year is typically greater than the percentage employed at a point in time. This comparability is in no way heartening, however. It means that half of the 17- to 19-year-old STEP youth who were not in school were not producing income, or gaining work habits or technical skills.

Parenting

Among Cohort III girls (17- and 18-year-olds), approximately 22 percent of the treatments and controls had a child. Among Cohort II girls (18- and 19-year-olds), approximately 33 percent had children. Among Cohort III mothers, approximately 30 percent were still in high school and an additional 20 to 25 percent were enrolled in some type of non-high school education. These are smaller percentages than are found in the general Cohort III sample--of whom about 50 percent were still in high school and approximately 20 percent were pursuing some sort of non-high school education. Among the older Cohort II mothers, approximately 7 percent were still in high school, compared with 12 percent of the overall Cohort II sample. Other educational opportunities were being pursued by an additional 35 percent of the mothers, compared with about 40 percent among the general Cohort II sample.

The majority of the girls who were mothers received public assistance in some form, with higher participation by the older mothers. Sixty-four percent (64%) of the younger Cohort III mothers and 75 percent of the older Cohort II mothers had received WIC or AFDC payments at some time since they became pregnant. Seventy percent (70%) of the pregnant girls in both cohorts had used WIC, and, at the time of the interview, 56 percent of Cohort III mothers and 68 percent of Cohort II mothers were receiving AFDC payments. It may be that more of the younger mothers were covered under their own mothers' cases, while the older mothers had cases on their own.

In sum, a substantial share of the STEP population, both controls and treatments, was not working, not in school, or had become parents without the personal or family means to support their children. The older Cohort II shows higher percentages in all these categories, raising the possibility that the Cohort III rates will rise further. At whatever level these rates stabilize, they are alarmingly high.

Why No Long-Term Impacts?

The STEP demonstration tells us clearly and conclusively what does **not** produce long-term changes in the school and work lives of 14- and 15-year-old poor, educationally disadvantaged youth. The 1990-91 long-term data show what the natural status of STEP youth is at ages 17 to 19, and the serious difficulties many of them will have reversing their life course.

The preciseness of the STEP model, the consistency and fidelity of its implementation, and the consistency of both its short- and long-term impacts, provide confidence that the lack of long-term impacts arises from inadequacy in the STEP intervention theory and its implementation model, not from poor or inconsistent execution of the model or the research design.

Although that confidence allows us to rule out major possible reasons for STEP's failure, it does not allow us to pinpoint exactly which shortcomings in the intervention theory or its implementation model are the critical ones. However, the serious condition of STEP youth in 1990-91 makes it unlikely that modest improvements in the STEP approach would alter the results.

For many youth, STEP was but a brief positive point in an otherwise declining life course **vis-a-vis** economic self-sufficiency.

Thus, it is important that we articulate the **major** areas of STEP's theoretical and model weakness. Our reflections on STEP, and our examination of adolescent development and education theory and research, lead us to the following conclusions:

First, STEP was not in fact preventive for this population. Most STEP youth, though only 14 and 15, were already doing poorly in school. For example, virtually all STEP youth were reading below grade and almost one-third had failed at least one grade. National statistics, and the 1990-91 data regarding the status of STEP youth, indicate how difficult it is to alter the life course of a young person who has few family resources, and who has already had formal signals of poor school performance, such as grade failure.

This lack of preventiveness is crucial because STEP was a quantitatively and qualitatively modest intervention. The intervention theory and the model assumed that a "boost" was all STEP youth needed--as if they had been targeted for intervention before serious problems had emerged in their lives. This assumption rested on the fact that the youth were 14 and 15--younger than most special interventions had previously targeted--and that they had not yet dropped out. But the assumption turned out to underrate the extent and importance of their problems; how close ages 14 and 15 are to dropping out and parenting decisions; and the extent to which those decisions are the results of many experiences and influences, past and current. A modest, brief intervention was no match for life and school performance problems that were already well-formed.

Second, the model had weak or nonexistent reinforcement mechanisms to connect the summer experience to the school year, or to other key aspects of their lives. STEP youth--as evidenced by their attendance, reenrollment rates, test scores, and attitudes about work experience and the LSO classes--did respond successfully to STEP. But STEP was self-contained: neither parents nor adult mentors were involved; the brevity of the summer, and the scattering of youth to different high schools and classes, did not allow for peer group formation; and STEP and school-year teachers and guidance counselors had no joint sessions to discuss the strengths, weaknesses, and learning and life issues of individual youth. Also, school-year support played a strong or reinforcing role in the lives of only a very few youth.

None of these reinforcing mechanisms are easy to implement, as other initiatives have evidenced. Without them, however, there seems little chance that STEP or other self-improvement strategies can succeed, especially if they deal with youth who are already having significant problems.

Major self-improvement or life change strategies--ranging from Alcoholics Anonymous to highly structured religious or political movements--all rely on planned reinforcement to ensure that their desired changes hold. To attempt otherwise in youth self-improvement programs defies the evidence from a massive body of experience and theory about life course redirection.

Third, STEP's design assumed that improved school-year performance would be primarily a product of the learning gains and a boost in confidence that could be achieved by successful learning over the summer, and had little to do with the structure, content or approach by which learning took place in

the school year. But the STEP *Practical Academics* and *Life Skills* curricula were very different in design and execution from the content and learning approaches of the high schools to which youth returned. In addition, they were accompanied by paid work. What worked in STEP was very different from what took place in school. Of the self-reported dropouts in Cohorts II and III, about 30 percent said they dropped out because they simply did not like regular school. If structure, content and approach are meaningful components of learning--as much modern learning theory and evidence indicates--then there was simply too little of STEP to make a long-term difference.

Fourth, STEP did not seek to exert any major environmental influence. It not only had no influence on and little interaction with the schools, it also had no interaction with or influence on any of the other major influences in a youth's life, including peers, neighborhood, family, family income, and perceived and real future job opportunities. Affecting such environmental factors is a tall order for a special program initiative like STEP, but that does not diminish the importance of those factors. For example, almost half of the STEP dropouts cited the need for income or other causes in their environment as their primary reason for dropping out of school.

As noted earlier, STEP's operation and results do not point clearly at any one factor as the key to achieving long-term results. But they do, in our judgment, push for a harder look at major issues. STEP has found a place in many communities as an effective summer program, but neither STEP nor similar brief, self-improvement programs are likely to alter the portrait of STEP-like youth in the '90s. To achieve redirection in the life course of a poor young teenager already having problems will, in our judgment, require at a minimum an intervention with more sustained length and powerful reinforcements than STEP. Given the success of STEP innovations, and the need for duration and reinforcement, it will probably require changes in the ways schools teach and involve youth. The data on parenting and working suggest that many youth may also need a more structured connection to the labor market, and a stronger sense of possible careers and their benefits.

Thus, our reflection on STEP's lack of impacts is sobering; it does not lead us to think that tinkering with STEP will achieve long-term impacts. More fundamental improvements are needed. But it is encouraging that, through demonstrations like STEP, we are coming closer to defining what elements, either singly or together, must be put in place to improve the prospects of a large share of our country's children.

VI. Thoughts For The Future

The STEP experience, comprising both a rigorous research demonstration over an eight-year period and a sizable replication effort, is rich in lessons and insights as well as hard data. It was operated in a large number of varied settings, and touched on many of the issues relevant to any effort to assist youth--including institutional coordination, the development and installation of innovative curricula and teaching methods, and multicomponent programming. STEP is therefore worth mining for all it can tell us about the shape and content of future initiatives.

The current status of poor youth in America makes this mining especially important. Their numbers and shares are rising; their environments are steadily declining in supports, resources and opportunities. Their economic opportunities and achievements as adults are likewise declining. Thus, the need for information and exploration of how these trends might be reversed is critical.

This chapter briefly describes the major lessons and insights we have derived from the STEP demonstration and replication experience. They are organized under the three broad questions posed at the beginning of the report.

1. **Does STEP advance our understanding of what is required to substantially reduce serious social problems, such as school dropout and teenage pregnancy**? At the conclusion of Chapter V, we noted several factors--prevention, reinforcement, learning theory, environmental influence and perceived opportunities--that were not, on reflection, adequately treated in STEP and, in some combination, may account for the lack of long-term impacts. Although STEP cannot tell us exactly how, or in what combination, those factors can be used to change youth's lives, they

do constitute guideposts that can help shape new initiatives, and the expectations we have for them.

But perhaps more important than STEP's indication of the factors that need more careful thought, is its implication that we need well-developed theories of how humans change. Our intervention strategies are currently most often based on youthful needs and adult hopes--and those, as documented in STEP's development and results, are many. Program strategies based on needs and hopes, however, do not constitute theories about redirection and change. Yet it is redirection and change, over a lifetime, that most special interventions like STEP seek.

STEP did not have a well-developed, long-term change theory. It did have a sound theory on how to achieve short-term improvements, then **assumed** that further changes would be forthcoming regardless of the major environmental or institutional influences that continued to play major roles in the youth's lives. In future initiatives, that area of assumption must be fleshed out. The insights and data generated from motivational psychology, adolescent development, learning theory and environmental influence studies, as well as the experience of other efforts engaged in inducing long-term change in other aspects of life, should provide useful building material for the development of soundly-conceived theories.

More explicit change theories should also be helpful in establishing appropriate expectations for particular youth services and programs. For example, STEP's designers and other researchers identified the summer as a major period of educational decline in a poor youth's life, and as a period for which no major institution in a poor community (e.g., school or family) takes responsibility for structuring. The desire to fill that major gap productively appears correct, even in hindsight. That filling it does not produce long-term impacts may not mean that nothing should be done with the summer period; a better-developed change theory may see that as a necessary but insufficient element of long-term change. The same line of reasoning applies to other important gap areas, such as after school, weekends and school-to-work transition.

In short, better developed change theory would not only allow us to fashion initiatives with a higher likelihood of long-term impacts; it would also permit us to construct appropriate expectations for programs like STEP.

STEP also suggests a more concrete lesson about reducing serious social problems like school dropout and teen pregnancy, i.e., that such reduction cannot be achieved without substantive involvement of such major natural influences as school, community organizations, peers, family and labor market. STEP operated in isolation from those natural influences; its impacts evaporated quickly. If the main influences on youth cannot be utilized in a constructive manner, it is difficult to see how young people can be induced to change their life trajectory.

Finally, STEP corroborated the importance and cost of positive adult attention and contact in achieving youthful progress. The *Practical Academics* and *LSO* classes are not

only innovative in content, but are administered at a 1:15 teacher/student ratio, and include time for individualized instruction. Demonstration operators attracted large numbers of students for STEP's second summer only by means of individualized outreach and contact, often including home visits.

Neither the schools nor employment training agencies in the STEP demonstration had sufficient resources to implement this level of adult contact at scale. Supplementary funding was provided in the demonstration to achieve the adult contact required in STEP. This last point should help establish a bottom line for the various efforts at school reform, and at improvement in other youth service organizations: if they cannot produce more positive adult contact in various forms, they may not be worth the effort, for either short- or long-term impacts.

2. **Does STEP suggest changes in the way research demonstrations are carried out, in what we expect from them or how we use them**? The need for more explicit and well-grounded change theories is particularly important to youth research demonstrations. Demonstrations represent a rare opportunity to transform the real world into a laboratory; in addition, the credibility of their findings makes it vital that they test good theory. The research demonstration is the best-suited vehicle for withstanding the pressures and compromises that reduce the knowledge-producing value of most social initiatives.

The need for more explicit theories in research demonstrations represents a conscious evolution of current practice. Early demonstrations could, with some justification, note that there was so little basis for theory that it made sense just to test something, and begin building theory as facts emerged.

But over the past two decades, there has been an enormous amount of academic work in fields like adolescent development, behavior modification and learning theory. In addition, youth demonstrations themselves have produced a substantial body of knowledge that can be used to generate theory. There is now less justification for proceeding to a research demonstration without an explicit, grounded theory based on accepted principles of human development. The time, expense and attention paid to demonstration findings also generates the responsibility to ensure that each demonstration has a credible theory of change.

Demonstrations are used primarily to learn about long-term impacts. STEP suggests the usefulness of the research demonstration approach for generating and measuring short-term outcomes. If development experts are correct that youth need a number and variety of positive experiences to mature successfully--and that those experiences should occur in various areas and times of their lives, including the "gap" areas mentioned earlier--then it is important that we know what to expect from individual interventions (like STEP), and how to measure their immediate effects. If our measure of each youth program and intervention is solely its ability to generate long-term impacts, we will probably end up concluding that no one program or intervention is important to a youth's life; whereas each should be assessed for its ability to contribute to successful maturation.

The research demonstration approach is a good one for developing and testing short-term developmental effects. Its emphasis on clarity of purpose and strict evidentiary standards works well on defined, time-limited activities. The challenge is to utilize new knowledge about adolescent development to shape credible measures of different aspects of the maturation process.

The STEP experience also highlights the use of demonstrations in generating adequate investment in social innovations and their installation. STEP created new approaches to teaching educational and life skills, and produced curricula and guides to aid in implementing those new approaches. The project had adequate resources for developing those innovations, and in training local staff in their use. In our judgment, it was mainly the sizeable investment in substantive innovation and local staff training that brought about the consistency of implementation quality and positive short-term impacts across the nation and among different subgroups that has been so rarely achieved in national social programs.

Given the public sector's generally low level of investment in innovations and training, the research demonstration's established practice of securing a combination of philanthropic, corporate and public funds makes it a useful device for developing, testing and installing complex innovations. The resulting knowledge about installation practice, costs and effects that demonstrations like STEP produce--as well as the production of the innovations themselves--should in this period of limited public resources become a major purpose and product of the demonstration process, rather than an unmined and underused byproduct.

In sum, the STEP experience highlights the innovation investment function of research demonstrations; the need to develop stronger conceptual theories of change; and the advisability of focusing on short-term developmental impacts. Observation of these lessons would, in our judgment, increase the research demonstration's usefulness for youth policy and practice.

3. **Does STEP's replication experience inform us about the development and operation of large-scale social programs**? STEP's replication in more than 100 localities is unusual for research demonstration interventions. The existence of STEP's impressive summer-learning gains, the JTPA amendment requiring remediation for youth in the summer jobs program, and the availability of additional funding were the key factors in making the replication possible.

   The financing of the replication was more complex than the demonstration itself. State governments and private corporations were added to the demonstration mix of philanthropies and the federal government. The philanthropic and corporate sectors provided financing for the development and marketing of STEP to new locations; states supplemented private and federal funds in covering training and installation costs; localities assumed the incremental costs of STEP programming and materials.

The primary reason for the complex financing package is that large-scale replication is not an institutionalized tool of the social policy world, especially for interventions like STEP that cross institutional or policy boundaries (in STEP's case, education and employment training). Yet it is boundary-crossing that must occur if we are to broaden and change the services available to youth. Thus, the basic lesson from STEP replication is how little experience and knowledge is available among public funding sources on how their funds can work together to mount a program that cuts across their institutional boundaries. In addition, several of the key functions of successful replication--marketing, materials packaging and quality control systems, for example--are not easily covered by the typical budget allocations of public agencies, and had to be covered by corporate and philanthropic sources.

This lesson does not portend well for widespread innovative and coordinated programming unless a stronger central investment, training and oversight role is assumed. Legislative mandate and exhortation are inadequate.

The second major lesson is more hopeful: there was no magic to large-scale replication of a complex social program and its achievement of consistent results. Resources, materials, skilled training of staff and high levels of quality control were the major requirements.

The need for an "exceptional and unique leader," which is often represented as the major obstacle to the spread of innovative programs with consistent performance, was not important to effective local implementation. Our guess is that the "unique leader" syndrome would have been more significant had the educational components, financing package and staff training been left entirely to each locality. In STEP's case, those functions were handled nationally, and we found effective and competent local implementation staff to be widely available.

Our STEP replication work also leads us to suspect that the oft-cited dichotomy between "top-down" and "bottom-up" programming--or between "cookie-cutter" and "local creativity" approaches--signals a debate over the wrong issue. The STEP program, for example, represented a balance between central and local functions that respected a local desire for structure and specific technical assistance, as well as for autonomy.

The issue seems less a matter of choosing between "top-down" and "bottom-up" than in assigning the various functions involved in mounting a new program initiative to the level or organizations where appropriate staff and funding resources are available to carry them out best, and in establishing a process in which there is regular discussion and feedback among the various levels on the performance of their functions, and how performance can be improved. As long as local/state/national relationships in social programming revolve around slogans, it is difficult to see how complex partnerships like those achieved in STEP can become representative, and not exceptional.

The usual training approach in the social policy world emphasizes the dissemination of general principles and broad lessons, leaving it to local staff to utilize these principles and lessons in their own initiatives. The STEP replication experience suggests that the experiential, hands-on learning approach that

has recently gained so much currency in educational circles also has value as a staff development tool. Learning concretely how to do one program has overall staff development benefits that rival those of absorbing general principles and broad lessons. The number of anecdotes and testimonials concerning staff perceptions of their own professional growth through the operation of STEP raised for us the issue of replication's potential as a staff development tool, an additional avenue for exploration.

* * *

STEP was a highly defined, time-limited intervention targeting low-income 14- to 15-year-olds with academic deficiencies, operated primarily by employment training agencies with various levels of cooperation from the public schools. Its experience and findings are, strictly speaking, limited to those circumstances. But its lessons have provoked us to broader thoughts about how to improve youth's prospects, how to use the research demonstration process in so doing, and how to improve the functioning and capacity of youth-serving institutions. Those broader insights and speculations do not cohere into a concrete strategy for action, but they represent new directions for funders, policymakers and administrators to reflect on as they shape their plans for the future.

References

Angrist, Joshua and Alan Krueger. "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics*, November 1990.

Berlin, Gordon. *Remedial Education, Social Competence and Work Experience for Disadvantaged Youth: Proposal for a Redesign of the Summer Youth Employment Program*. Unpublished Manuscript, October 1983.

Berlin, Gordon, and Andrew Sum. *Toward a More Perfect Union: Basic Skills, Poor Families, and Our Economic Future*. Occasional Paper 3, Ford Foundation Project on Social Welfare and the American Future. New York: The Ford Foundation, 1988.

Betsey, Charles L., Robinson G. Hollister, Jr. and Mary R. Papageorgiou, eds. *Youth Employment and Training Programs: The YEDPA Years*. Committee on Youth Employment Programs, Commission on Behavioral and Social Sciences and Education, National Research Council. Washington, D.C.: National Academy Press, 1985.

Branch, Alvia Y., Julita Milliner and Jon Bumbaugh. *Summer Training and Education Program (STEP): Report on the 1985 Summer Experience*. Philadelphia: Public/Private Ventures, April 1986.

Branch, Alvia Y., Julita Milliner, Susan Phillipson Bloom and Jon Bumbaugh. *Summer Training and Education Program: Report on the Pilot Experience*. Philadelphia: Public/Private Ventures, March 1985.

Card, David and Alan Krueger. "Does School Quality Matter? Returns to Education and Characteristics of Public Schools in the United States." *Journal of Political Economy*, February 1992.

Carter, Deborah and Reginald Wilson. *Minorities in Higher Education. Eighth Annual Status Report*. Washington, D.C.: American Council on Education, Office of Minority Concern, 1989.

Cave, George, and Janet Quint. *Career Beginnings Impact Evaluation: Findings From a Program for Disadvantaged High School Students*. New York: Manpower Demonstration Research Corporation, 1990.

Coombs, J., and W. W. Cooley. "Dropouts: In High School and After High School," *American Educational Research Journal*, 1968, Vol. 5:343-358.

David, Jane, and Sol H. Pelavin. "Secondary Analysis: In Compensatory Education Programs," *New Directions for Program Evaluation*, 1978, Vol. 4:31-44.

De Lone, Richard H. *Replication: A Strategy to Improve the Delivery of Education and Job Training Programs*, Philadelphia: Public/Private Ventures, Summer 1990.

Dryfoos, J. G. *Prevention Strategies: A Progress Report*. Report to the Rockefeller Foundation, New York, 1985.

Griliches, Z. and William Mason. "Education, Income and Ability," *Journal of Political Economy*, May/June 1972, Vol. 80:2, S74-S103.

Grossman, Jean Baldwin, and Cynthia L. Sipe. *Summer Training and Education Program (STEP): Report on Long-Term Impacts*, Philadelphia: Public/Private Ventures, Winter 1992.

Gueron, Judy. *Lessons from a Job Guarantee: The Youth Incentive Entitlement Pilot Projects*. New York: Manpower Demonstration Research Corporation, 1984.

Hammond, Pierce A. III and Joy A. Frechtling. "Twelve, Nine and Three Month Achievement gains of Low and Average Achieving Elementary School Students." Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1979.

Hayes, Cheryl D., ed. *Risking the Future: Adolescent Sexuality, Pregnancy and Child Bearing*. Report of the Panel on Adolescent Pregnancy and Childbearing, National Research Council. Washington, D.C.: National Academy Press, 1987.

Hayes, Donald P., and Judith Grether. "The School Year and Vacations: When Do Students Learn?" Revision of a paper presented at the Eastern Sociological Association Convention, New York City, April 1969.

Heyns, Barbara. *Summer Learning and the Effects of Schooling*. New York: Academic Press, 1978.

Heyns, Barbara. "Schooling and Cognitive Development: Is There a Season for Learning?" *Child Development*, 1987, Vol. 58: 1151-60.

Hollister, Robinson G., Jr., Peter Kemper and Rebecca Maynard, eds. *The National Supported Work Demonstration*, Madison: The University of Wisconsin Press, 1984.

Hungerfrond, Thomas and Gary Solon. "Sheepskin Effects in the Returns to Education," *Review of Economics and Statistics* Vol. 66, February 1987, pp 175-77.

Institute for Educational Leadership, Inc., *School Dropouts: Everybody's Problem*. Report of the Institute for Educational Leadership Conference, Washington, D.C., 1986.

Littell, Julia and Joan Wyna. *The Availability and Use of Community Resources for young Adolescents in an Inner City and Suburban Community*, Chicago: The Chapin Hall Center for Children at the University of Chicago, 1989.

Meyer, Robert H. and David A. Wise. "High School Preparation and Early Labor Force Experience." *The Youth Employment Problem: Dimensions, Causes and Consequences*. Washington, D.C.: National Bureau of Economic Research, 1980.

Mincer, J. *Schooling, Experience and Earnings*, New York: National Bureau of Economic Research, 1974.

National Center for Education Statistics, U.S. Department of Education. *Dropout Rates in the United States: 1989*, Office of Education Research and Improvement (NCES 90-659), Washington, D.C.: 1990.

National Center for Education Statistics, U.S. Department of Education. *The Condition of Education*, 1985 Edition. Washington, D.C.: U.S. Government Printing Office, 1985.

Natriello, Gary. *School Dropouts: Patterns and Policies*. New York: Teachers College Press, 1987.

Pelavin, Sol H., and Jane David. *Evaluating Long-Term Achievement: An Analysis of Longitudinal Data From Compensatory Education Programs*. Research Report, SRI Project 4537. Menlo Park, California: Stanford Research Institute, International, March 1977.

Presser, H.B. "Social Consequences of Teenage Childbearing," in C. Chilman (ed.), *Teenage Childbearing: Recent Research on the Determinants and Consequences*. Washington, D.C.: U.S. Government Printing Office, 1978.

Quint, Janet C. and James A. Riccio. *The Challenge of Serving Pregnant and Parenting Teens: Lessons from Project Redirection*. New York: Manpower Demonstration Research Corporation, April 1985.

Simons, Janet M., Belva Finley and Alice Yang. *The Adolescent and Young Adult Fact Book*, Washington, D.C.: The Children's Defense Fund, 1991.

Sipe, Cynthia L., Jean Baldwin Grossman and Julita A. Milliner. *Summer Training and Education Program (STEP): Report on the 1986 Experience*. Philadelphia: Public/Private Ventures, April 1987.

Sipe, Cynthia L., Jean Baldwin Grossman and Julita A. Milliner. *Summer Training and Education Program (STEP): Report on the 1987 Experience*. Philadelphia: Public/Private Ventures, Summer 1988.

Sum, Andrew, et al. *Educational Attainment, Academic Ability, and the Employability and Earnings of Young Persons: Implications for the Planning and Design of JTPA Youth Programs*. Prepared for a Symposium on Serving Disadvantaged School Dropouts. Waltham, Massachusetts: Brandeis University, December 1983.

Taggart, Robert. *A Fisherman's Guide: An Assessment of Training and Remediation Strategies*. Kalamazoo, Michigan: W.E. Upjohn Institute, 1981.

U.S. Department of Commerce, Bureau of the Census. *Lifetime Earnings Estimates for Men and Women in the United States*, Current Population Reports, Series p. 60, No. 139. Washington, D.C.: U.S. Government Printing Office, 1983.

Weis, L., E. Farrar and H. Petrie. *Dropouts From School: Issues, Dilemmas, and Solutions*. Albany, New York: State University of New York Press, 1989.

Willis, Robert. "Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions," *Handbook of Labor Economics* edited by O. Ashenfelter and B. Layard. Amsterdam: Elsevier Science Publishers BV, 1986.

Public/Private Ventures
399 Market Street
Philadelphia, PA 19106

Winter 1992